CHAPTER VIII

# QUADRATIC RISK OF LINEAR REGRESSION WITH A LARGE NUMBER OF RANDOM PREDICTORS

$1^o$ We consider a class of generalized regularized sample regression procedures depending on an arbitrary function.

$2^o$ We single out the leading part of the increasing dimension asymptotics for the quadratic risk of linear regressions with random predictors and obtain upper estimates for the remainder terms.

$3^o$ We construct an estimator for the leading part of the quadratic risk consistent uniformly with respect to distributions.

Suppose that an $(n+1)$-dimensional population $\mathfrak{S}$ is given in which the observations are pairs $(\mathbf{x}, y)$, where $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of predictors and $y$ is a scalar response.

Define the centered values $\overset{o}{\mathbf{x}} = \mathbf{x} - \mathbf{E}\ \mathbf{x}$, and $\overset{o}{y} = y - \mathbf{E}\ y$. We restrict the population by the requirement that the four moments of all predictors exist and there exists the moment $M_8 = \mathbf{E}\ (\overset{o}{\mathbf{x}}{}^2/n)^2 \overset{o}{y}{}^4$ (here and in the following, squares of vectors denote the squares of lengths). Assume, additionally, that $\mathbf{E}\, \overset{o}{\mathbf{x}}{}^2 > 0$ (non-degenerate case). Denote

$$M_4 = \sup_{|\mathbf{e}|=1} \mathbf{E}\ (\mathbf{e}^T \overset{o}{\mathbf{x}})^4 > 0, \quad M = \max\ (M_4,\ \sqrt{M_8},\ \mathbf{E}\ \overset{o}{y}{}^4),$$

$$\text{and} \quad \gamma = \sup_{\|\Omega\|=1} \text{var}\ (\overset{o}{\mathbf{x}}{}^T \Omega \overset{o}{\mathbf{x}}/n)/M, \tag{1}$$

where (and in the following) $\mathbf{e}$ is a non-random vector of unit length, and $\Omega$ is a symmetric positive semidefinite matrix with unit spectral norm. We consider the linear regression $y = \mathbf{k}^T\mathbf{x} + l + \Delta$, where $\mathbf{k} \in \mathbb{R}^n$ and $l \in \mathbb{R}^1$. The problem is to minimize the quadratic risk

$R = \mathbf{E}\ \Delta^2$ by the best choice of $\mathbf{k}$ and $l$ that are calculated over a sample $\mathfrak{X} = \{(\mathbf{x}_m, y_m)\}$, $m = 1, \ldots, N$, from $\mathfrak{S}$.

We denote $\lambda = n/N$, $\mathbf{a} = \mathbf{E}\ \mathbf{x}$, $a_0 = \mathbf{E}\ y$, $\Sigma = \text{cov}\ (\mathbf{x}, \mathbf{x})$, $\sigma^2 = \text{var}\ y$, and $\mathbf{g} = \text{cov}\ (\mathbf{x}, y)$.

If $\sigma > 0$ and the matrix $\Sigma$ is non-degenerate, then the a priori coefficients $\mathbf{k} = \Sigma^{-1}\mathbf{g}$ and $l = a_0 - \mathbf{k}^T\mathbf{g}$ provide a minimum of $R$, which is $R = \sigma^2 - \mathbf{g}^T\Sigma^{-1}\mathbf{g} = \sigma^2(1 - r^2)$, where $r$ is the multiple correlation coefficient.

We start from the statistics

$$\bar{\mathbf{x}} = N^{-1}\sum_{m=1}^{N}\mathbf{x}_m, \quad \bar{y} = N^{-1}\sum_{m=1}^{N} y_m, \quad \widehat{\sigma}^2 = N^{-1}\sum_{m=1}^{N}(y_m - \bar{y})^2,$$

$$S = N^{-1}\sum_{m=1}^{N}\mathbf{x}_m\mathbf{x}_m^T, \qquad \widehat{\mathbf{g}}_0 = N^{-1}\sum_{m=1}\mathbf{x}_m y_m,$$

$$C = N^{-1}\sum_{m=1}^{N}(\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T, \ \widehat{\mathbf{g}} = N^{-1}\sum_{m=1}^{N}(\mathbf{x}_m - \bar{\mathbf{x}})(y_m - \bar{y}).$$

The standard 'plug-in' procedure with $\mathbf{k} = C^{-1}\widehat{\mathbf{g}}$ and $\widehat{l} = \bar{y} - \mathbf{k}^T\bar{\mathbf{x}}$ has known demerits: this procedure does not guarantee the minimum risk, is degenerate for multi-collinear data (for a degenerate matrix $C$), and is not uniformly consistent with respect to the dimension.

The quadratic risk of the regression $y = \widehat{\mathbf{k}}^T\mathbf{x} + \widehat{l} + \Delta$, where $\widehat{\mathbf{k}}$ and $\widehat{l}$ are calculated over a sample with the 'plug-in' constant term $\widehat{l} = \bar{y} - \widehat{\mathbf{k}}^T\bar{\mathbf{x}}$, is given by

$$R = \mathbf{E}\ \Delta^2 = R^1 + \mathbf{E}\ (\bar{y} - \widehat{\mathbf{k}}^T\bar{\mathbf{x}})^2 = (1 + 1/N)R^1,$$

where

$$R^1 \stackrel{\text{def}}{=} \mathbf{E}\ (\sigma^2 - 2\widehat{\mathbf{k}}^T\mathbf{g} + \widehat{\mathbf{k}}^T\Sigma\widehat{\mathbf{k}}). \tag{2}$$

Let us calculate and minimize $R^1$. We consider the following class of generalized regularized regressions. Let $H_0 = (I + tS)^{-1}$ and $H = (I + tC)^{-1}$ be the resolvents of the matrices $S$ and $C$, respectively.

We choose the coefficient $\widehat{\mathbf{k}}$ (everywhere below) in the class $\mathfrak{K}$ of statistics of the form $\widehat{\mathbf{k}} = \Gamma\widehat{\mathbf{g}}$, where

$$\Gamma = \Gamma(t) = \int tH(t)d\eta(t)$$

and $\eta(t)$ are functions whose variation on $[0, \infty)$ is at most one and that has sufficiently many moments $\eta_k \stackrel{\text{def}}{=} \int t^k |d\eta(t)|, \quad k = 1, 2, \ldots$ The function $\eta(t)$ formed by a unit jump corresponds to the 'ridge regression' [1]. The regression with the coefficients $\widehat{\mathbf{k}} \in \mathfrak{K}$ can be called a generalized ridge regression. The quantity (2) depends on $\eta(t)$, $R^1 = R^1(\eta)$, and

$$R^1(\eta) = \sigma^2 - 2\mathbf{E} \int t\mathbf{g}^T H(t)\widehat{\mathbf{g}} d\eta(t) + \iint D(t, u) d\eta(t) d\eta(u), \quad (3)$$

where

$$D(t, u) \stackrel{\text{def}}{=} tu\widehat{\mathbf{g}}^T H(t) \Sigma H(u) \widehat{\mathbf{g}}.$$

Since all arguments of $R^1(\eta)$ are invariant with respect to the translation of the origin, we assume (everywhere in the following) that $\mathbf{a} = \mathbf{E}\ \mathbf{x} = 0$ and $a_0 = \mathbf{E}\ y = 0$.

Our purpose is to single out leading parts of these functionals and obtain upper bounds for the remainder terms up to absolute constants. To simplify notations of the remainder terms, we write

$$\tau = \sqrt{M} t, \quad \varepsilon = \sqrt{\gamma + 1/N},$$
$$c_{lm} = c_{lm}(t) = a \max\,(1, \tau^l) \max\,(1, \lambda^m),$$

where $a, l$ and $m$ are non-negative numbers (for brevity, we omit the parentheses in $c_{lm}(t)$ indicating the dependence on $t$). Starting from (1), we can readily see that

$$\mathbf{E}\ (\mathbf{x}^2)^2 \le M, \quad \mathbf{E}\ (\bar{\mathbf{x}}^2)^2 \le M\lambda^2, \quad \|\Sigma\|^2 \le M, \quad \mathbf{g}^2 \le M.$$

As in previous chapters, we begin by studying functions of more simple covariance matrices $S$ and then pass to functions of $C$.

### Spectral Functions of Sample Covariance Matrices

Our investigation will be based on results of the spectral theory of large sample covariance matrices developed in Chapter 2. To cite these results in a more convenient form, we restrict ourselves with real non-positive values of the complex argument $z$ of spectral functions so that $t = -z \geq 0$ and preserve the notations of functions. Thus, we define

$$H_0 = H_0(t) = (I + tS)^{-1}, \quad \widehat{h}_0(t) = n^{-1}\text{tr } H_0(t),$$
$$h_0(t) = \mathbf{E}\ \widehat{h}_0(t), \quad s_0 = s_0(t) = 1 - \lambda + \lambda h_0(t),$$
$$V = V(t) = \mathbf{e}^T H_0(t)\bar{\mathbf{x}}, \quad \Phi = \Phi(t) = \bar{\mathbf{x}}^T H_0(t)\bar{\mathbf{x}}\,.$$

Also we define

$$H = H(t) = (I + tC)^{-1}, \quad \widehat{h}(t) = n^{-1}\text{tr } H(t),$$
$$h(t) = \mathbf{E}\ \widehat{h}(t), \quad s = s(t) = 1 - \lambda + \lambda h(t),$$
$$U = U(t) = \mathbf{e}^T H(t)\bar{\mathbf{x}}, \quad \Psi = \Psi(t) = \bar{\mathbf{x}}^T H(t)\bar{\mathbf{x}}.$$

We formulate some results of Chapter 2 in the form of a lemma.

LEMMA 8.1. (corollary of Lemmas 2.1–2.4 and Theorems 2.1, 2.2).

$1^o$ $s_0 = \mathbf{E}\ (1 - t\psi_1) \geq (1 + \tau y)^{-1}$,

$\quad$ $\text{var } (t\psi_1) \leq \delta \stackrel{\text{def}}{=} 2\tau^2\lambda^2(\gamma + \tau^2/N) \leq c_{42}\varepsilon$;

$2^o$ $\mathbf{E}\ H_0 = (I + ts_0\Sigma)^{-1} + \Omega_0$, *where* $\|\Omega_0\| \leq c_{31}\varepsilon$,

$\quad$ $\text{var } (\mathbf{e}^T H_0 \mathbf{e}) \leq \tau^2/N$;

$3^o$ $t\ (\mathbf{E}\ V)^2 \leq c_{52}\varepsilon^2$, $\quad \text{var } (tV) \leq c_{20}/N$;

$4^o$ $t\Phi \leq 1$, $\quad t\ \mathbf{E}\ \Phi = 1 - s_0 + o$, $\quad o^2 \leq c_{52}\varepsilon^2$, $\quad \text{var } (t\Phi) \leq c_{20}/N$;

$5^o$ $\|\mathbf{E}\ H - \mathbf{E}\ H_0\| \leq c_{74}\varepsilon^2$, $\quad |s - s_0| \leq c_{11}/N$;

$6^o$ $\mathbf{E}\ H = (I + ts\Sigma)^{-1} + \Omega$, *where* $\|\Omega\|^2 \leq c_{63}\varepsilon^2$;

$7^o$ $U = V + tU\Phi$, $\quad (1 + t\Psi)(1 - t\Phi) = 1$;

$8^o$ $ts_0^2\ (\mathbf{E}\ U)^2 \leq c_{63}\varepsilon^2$, $\quad t\ \text{var } U \leq c_{63}/N$;

$9^o$ $ts_0\ \mathbf{E}\ \Psi = 1 - s + o$, *where* $o^2 \leq c_{74}\varepsilon^2$,

$\quad$ $\text{var } (t\Psi) \leq c_{64}/N$.

### Functionals Depending on the Statistics $S$ and $\widehat{\mathbf{g}}_0$

We use the method of the alternating elimination of independent sample vectors. Eliminating one of the sample vectors, say, the vector $\mathbf{x}_1$, we denote

$$\bar{\mathbf{x}}^1 = \bar{\mathbf{x}}_1 - \mathbf{x}_1/N, \quad \widehat{\mathbf{g}}_0^1 = \widehat{\mathbf{g}}_0 - \mathbf{x}_1 y_1/N,$$

$$S^1 = S - \mathbf{x}_1\mathbf{x}_1^T/N \quad H_0^1 = H_0^1(t) = (I + tS^1)^{-1}.$$

These values do not depend on $\mathbf{x}_1$ and $y_1$. The identity holds

$$H_0 = H_0^1 - tH_0^1\mathbf{x}_1\mathbf{x}_1^T H_0/N. \tag{4}$$

Denote

$$v_1 = v_1(t) = \mathbf{e}^T H_0^1(t)\mathbf{x}_1, \quad u_1 = u_1(t) = \mathbf{e}^T H_0(t)\mathbf{x}_1,$$
$$\varphi_1 = \varphi_1(t) = \mathbf{x}_1^T H_0^1(t)\mathbf{x}_1/N, \quad \psi_1 = \psi_1(t) = \mathbf{x}_1^T H_0(t)\mathbf{x}_1/N. \tag{5}$$

We have the identities

$$u_1 = (1 - t\psi_1)v_1, \quad (1 + t\varphi_1)(1 - t\psi_1) = 1, \quad H_0\mathbf{x}_1 = (1 - t\psi_1)H_0^1\mathbf{x}_1. \tag{6}$$

Obviously, $0 \le t\psi_1 \le 1$. It can be readily seen that

$$1 - s_0 = t\mathbf{E}\ \psi_1 = t\mathbf{E}\ N^{-1}\mathrm{tr}\ (H_0 S), \quad (1 + \tau\lambda)^{-1} \le s_0 \le 1. \tag{7}$$

From (1) it follows that

$$\|H_0\| \le \|H_0^1\| \le 1, \quad \mathbf{E}\ u_1^4 \le \mathbf{E}\ v_1^4 \le M. \tag{8}$$

**Remark 1.**

$$\mathbf{E}\ \widehat{\mathbf{g}}_0^2 \le M(1+\lambda), \qquad \mathbf{E}\ |\widehat{\mathbf{g}}_0|^4 \le 2M^2(1+\lambda)^2,$$
$$\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^2 \le M(1+\lambda), \qquad \mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4 \le 2M^2(1+\lambda)^2. \tag{9}$$

Indeed, the value $\widehat{\mathbf{g}}_0^1$ does not depend on $\mathbf{x}_1$ and $y_1$, and

$$\mathbf{E}\ \widehat{\mathbf{g}}_0^2 = \mathbf{E}\ y_1\mathbf{x}_1^T\widehat{\mathbf{g}}_0 = \mathbf{E}\ y_1\mathbf{x}_1^T\widehat{\mathbf{g}}_0^1 + \mathbf{E}\ y_1^2\mathbf{x}_1^2/N.$$

Here, the first summand equals $\mathbf{g}^2(1 - N^{-1}) \le M$. The second one is not greater $(\mathbf{E}\ y_1^4(\mathbf{x}_1^2/N)^2)^{1/2} \le M\lambda$. Further,

$$\mathbf{E}\ (\widehat{\mathbf{g}}_0^2)^2 = \mathbf{E}\ \widehat{\mathbf{g}}^2 y_1 \mathbf{x}_1^T \widehat{\mathbf{g}}_0 = \mathbf{E}\ \widehat{\mathbf{g}}_0^2 y_1 \mathbf{x}_1^T \widehat{\mathbf{g}}_0^1 + \mathbf{E}\ \widehat{\mathbf{g}}_0^2 y_1^2 \mathbf{x}_1^2/N.$$

Using the Schwarz inequality we find

$$\mathbf{E}\ (\widehat{\mathbf{g}}_0^2)^2 \le 2\mathbf{E}\ (y_1 \mathbf{x}_1^T \widehat{\mathbf{g}}_0^1)^2 + 2\mathbf{E}\ y_1^4 (\mathbf{x}_1^2)^2/N^2.$$

In the first summand here, for fixed $\widehat{\mathbf{g}}_0^1$, we have $\mathbf{E}\ y_1^2(\mathbf{x}_1^T \mathbf{e})^2 \le M$, where $\mathbf{e} = \widehat{\mathbf{g}}_0^1/|\widehat{\mathbf{g}}_0^1|$. Thus the first summand is not greater $2M\mathbf{E}\ \widehat{\mathbf{g}}_0^2 \le 2M^2(1+\lambda)$. By (1), the second summand is not greater $2M^2\lambda^2$. The second inequality in (9) follows. The same arguments establish the second pair of inequalities (9).

LEMMA 8.2. *If $t \ge 0$, then*

$$|t\mathbf{E}\ \bar{\mathbf{x}}^T H_0(t) \widehat{\mathbf{g}}_0| \le M^{1/4} c_{32} \varepsilon,$$
$$\mathrm{var}\ (t \bar{\mathbf{x}}^T H_0(t) \widehat{\mathbf{g}}_0) \le \sqrt{M} c_{42}/N.$$

Proof. Eliminating $(\mathbf{x}_1, y_1)$, we substitute $\widehat{\mathbf{g}}_0 = \tilde{\mathbf{g}}_0^1 + \mathbf{x}_1 y_1/N$. We have $\mathbf{E}\ \mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_1^1 = 0$, $\mathbf{E}\ y_1 = 0$. It follows

$$\begin{aligned} t\mathbf{E}\ \bar{\mathbf{x}}^T H_0 \widehat{\mathbf{g}}_0 &= t\mathbf{E}\ \mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0 = t\mathbf{E}\ \mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0^1 + \mathbf{E}\ t\psi_1 y_1 \\ &= t\mathbf{E}\ (1 - t\psi_1)\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1 + \mathbf{E}\ (1 - s_0 - \Delta_1) y_1 \\ &= t\mathbf{E}\ \Delta_1 \mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1 - \mathbf{E}\ \Delta_1 y_1, \end{aligned}$$

where $\Delta_1 = 1 - t\psi_1 - s_0$. By statement 1 of Lemma 8.1, $\mathbf{E}\ \Delta_1^2 \le \delta \le c_{42}\varepsilon^2$. Applying the Schwarz inequality, (1), and (9), we obtain that

$$\begin{aligned} (t\mathbf{E}\ \bar{\mathbf{x}}^T H_0 \widehat{\mathbf{g}}_0)^2 &\le [t^2 \mathbf{E}\ (\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1)^2 + \mathbf{E}\ y_1^2] c_{42} \varepsilon^2 \\ &\le \sqrt{M}\ [t^2 \mathbf{E}\ (\widehat{\mathbf{g}}_0^1)^2 + 1] c_{42} \varepsilon^2 \le \sqrt{M} c_{63} \varepsilon^2. \end{aligned}$$

The first statement follows.

To estimate the variance, we use Lemma 2.2. Let us eliminate the variables $\mathbf{x}_1$ and $y_1$. Denote $\tilde{\mathbf{x}} = \bar{\mathbf{x}} - \mathbf{x}_1/N$. Then $\bar{\mathbf{x}}^T H_0 \widehat{\mathbf{g}}_0$ is

$$\tilde{\mathbf{x}}^T H_0^1 \widehat{\mathbf{g}}_0^1 - t \tilde{\mathbf{x}}^T H_0^1 \mathbf{x}_1 \mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0^1/N + \mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0^1/N + y_1 \bar{\mathbf{x}}^T H_0 \mathbf{x}_1/N.$$

The first term in the right hand does not depend on $\mathbf{x}_1$ and $y_1$. In view of the identical dependence on sample vectors, we conclude that $t^2 \mathrm{var}\ (\bar{\mathbf{x}}^T H_0 \widehat{\mathbf{g}}_0)$ is not greater than

$$3\ [\mathbf{E}\ (t^2 \tilde{\mathbf{x}}_1^T H_0^1 \mathbf{x}_1 \mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0^1)^2 + \mathbf{E}\ (t\mathbf{x}_1^T H_0 \widehat{\mathbf{g}}_0^1)^2 + \mathbf{E}\ (t y_1 \bar{\mathbf{x}}^T H_0 \mathbf{x}_1)^2]/N.$$

In view of (6), this inequality remains valid if $H_0$ is replaced by $H_0^1$. After this replacement, we use (1). The square of the sum of the first two summand in the bracket is not greater than

$$\begin{aligned} &at^4 \mathbf{E}\ (\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1)^4 \mathbf{E}\ (t^2 (\tilde{\mathbf{x}}^T H_0^1 \mathbf{x}_1)^2 + 1)^2 \\ &\qquad\qquad \le aMt^4 \mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4 \mathbf{E}\ (\sqrt{M} t^2 \tilde{\mathbf{x}}_1^T (H_0^1)^2 \tilde{\mathbf{x}}_1 + 1), \end{aligned}$$

where $a$ is a numerical coefficient. From the definition, one can see that $t\Phi \le 1$ and $t\tilde{\mathbf{x}}_1^T (H_0^1)^2 \tilde{\mathbf{x}}_1 \le 1$. The square of the sum of the first two summands in the bracket is not greater than $aM^3 t^4 (\sqrt{M} t + 1)(1+\lambda)^2 \le M c_{52}$. In the third summand, $\bar{\mathbf{x}}^T H_0 \mathbf{x}_1 = \tilde{\mathbf{x}}^T H_0 \mathbf{x}_1 + \psi_1$. By (1) and (6), the square of this summand is not greater

$$2(t^4 \mathbf{E}\ y_1^4 \mathbf{E}\ (\tilde{\mathbf{x}}^T H_0^1 \mathbf{x}_1)^4 + M) \le 2M(Mt^4 \mathbf{E}\ |\tilde{\mathbf{x}}|^4 + 1) \le M c_{42}.$$

We conclude that the variance in the statement 4 is not greater $\sqrt{M} c_{31}/N$. The proof of Lemma 8.2 is complete. □

LEMMA 8.3.

$$t\mathbf{E}\ \mathbf{e}^T H_0(t) \widehat{\mathbf{g}}_0 = t s_0(t) \mathbf{E}\ \mathbf{e}^T H_0(t) \mathbf{g} + o,$$

$$\textit{where } |o| \le c_{31}\varepsilon; \qquad \mathrm{var}\ (t\mathbf{e}^T H_0(t) \widehat{\mathbf{g}}_0) \le c_{41}/N. \qquad (10)$$

Proof. Denote $\Delta_1 = t\psi_1 - t\mathbf{E}\ \psi_1$. Using (4) and (6), we find

$$\begin{aligned} t\mathbf{E}\ \mathbf{e}^T H_0 \widehat{\mathbf{g}}_0 &= t\mathbf{E}\ \mathbf{e}^T H_0 \mathbf{x}_1 y_1 = t\mathbf{E}\ (1 - t\psi_1) \mathbf{e}^T H_0^1 \mathbf{x}_1 y_1 \\ &= t s_0 \mathbf{E}\ \mathbf{e}^T H_0^1 \mathbf{x}_1 y_1 - t\mathbf{E}\ \Delta_1 \mathbf{e}^T H_0^1 \mathbf{x}_1 y_1 \\ &= t s_0 \mathbf{E}\ \mathbf{e}^T H_0^1 \mathbf{g} - t\mathbf{E}\ v_1 u_1 \Delta_1 \\ &= t s_0 \mathbf{E}\ \mathbf{e}^T H_0 \mathbf{g} + t^2 s_0 \mathbf{E}\ u_1 \mathbf{x}_1^T H_0^1 \mathbf{g}/N - t\mathbf{E}\ v_1 y_1 \Delta_1. \end{aligned}$$

The last two terms present the remainder term $o$ in the lemma formulation. We estimate these by the Schwarz inequality,

$$\begin{aligned} |o| &\le t^2 (\mathbf{E}\ u_1^2 \mathbf{E}\ (\mathbf{x}_1^T H_0 \mathbf{g})^2)^{1/2}/N + t\sqrt{\delta} (\mathbf{E}\ v_1^2 y_1^2)^{1/2} \\ &\le \tau^2/N + t\sqrt{\delta M} \le c_{31}\varepsilon. \end{aligned}$$

The first statement is proved.

Now we estimate the variance eliminating independent variables. Denote $f = t\mathbf{e}^T H_0 \widehat{\mathbf{g}}_0$. Let $f = f^1 + \Delta_1$, where $f^1$ does not depend on $\mathbf{x}_1$ and $y_1$. We have

$$f = \mathbf{e}^T H_0^1 \widehat{\mathbf{g}}_0^1 + t u_1 y_1/N + t^2 u_1 (\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1)/N.$$

By Lemma 2,2 we obtain var $f \le N\mathbf{E}\ \Delta_1^2$. Therefore,

$$\begin{aligned}\text{var } f &\le 2N^{-1}[t^4 \mathbf{E}\ u_1^2 (\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1)^2 + t^2 \mathbf{E}\ u_1^2 y_1^2] \\ &\le 2N^{-1} (\mathbf{E}\ u_1^4)^{1/2} M^{1/2} t^2 [t^2 (\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4)^{1/2} + 1] \\ &\le 2N^{-1} \tau^2 (2\tau^2 (1+\lambda) + 1) \le c_{41}/N.\end{aligned}$$

Lemma 8.3 is proved.□

LEMMA 8.4. *If $t \ge 0$, then*

$$\begin{aligned}t\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0 \widehat{\mathbf{g}}_0 &= \sigma^2 (1 - s_0) + t s_0 \mathbf{E}\ \mathbf{g}^T H_0 \widehat{\mathbf{g}}_0 + o_1 \\ &= \sigma^2 (1 - s_0) + t s_0^2 \mathbf{E}\ \mathbf{g}^T H_0 \mathbf{g} + o_2, \qquad (11)\end{aligned}$$

*where* $|o_1| \le \sqrt{M} c_{32} \varepsilon, \ \ |o_2| \le \sqrt{M} c_{32} \varepsilon$; $t^2 \text{var } (\widehat{\mathbf{g}}_0^T H_0 \widehat{\mathbf{g}}_0) \le M c_{42}/N$.

Proof. Using (4) and (6), we find that $\mathbf{E}\ t\widehat{\mathbf{g}}_0 H_0 \widehat{\mathbf{g}}_0 = \mathbf{E}\ t\mathbf{x}_1^T y_1 H_0 \widehat{\mathbf{g}}_0$ is equal to

$$\begin{aligned}&\mathbf{E}\ t\mathbf{x}_1^T y_1 H_0 \mathbf{x}_1 y_1/N + \mathbf{E}\ t\mathbf{x}_1^T y_1 H_0 \widehat{\mathbf{g}}_0^1 \\ &\qquad = \mathbf{E}\ t\psi_1 y_1^2 + \mathbf{E}\ t\mathbf{x}_1 y_1 (1 - t\psi_1) H_0^1 \widehat{\mathbf{g}}_0^1.\end{aligned}$$

Substituting $t\psi_1 = 1 - s_0 - \Delta_1$, we find that $\mathbf{E}\ t\widehat{\mathbf{g}}_0^T H_0 \widehat{\mathbf{g}}_0$ is

$$\sigma^2 (1 - s_0) - \mathbf{E}\ \Delta_1 y_1^2 + t s_0 \mathbf{E}\ y_1 \mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1 + t\mathbf{E}\ \Delta_1 y_1 \mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1.$$

The square of the second term in the right hand side is not greater than $M\delta \le M c_{42} \varepsilon^2$. Using the Schwarz inequality we obtain that the square of the fourth term is not greater than

$$\begin{aligned}&\delta t^2 \mathbf{E}\ y_1^2 (\mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1)^2 \\ &\qquad \le \delta M t^2 (\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4)^{1/2} \le 2\delta M \tau^2 (1+\lambda) \le M c_{63} \varepsilon^2.\end{aligned}$$

In the third term, $\mathbf{E}\ y_1\mathbf{x}_1^T = \mathbf{g}^T$, and by (4), we find that the third term is

$$ts_0\mathbf{E}\ \mathbf{g}^T H_0^1\widehat{\mathbf{g}}_0^1 = ts_0\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0^1 + t^2 s_0\mathbf{E}\ \mathbf{g}^T H_0\mathbf{x}_1\mathbf{x}_1^T H_0^1\widehat{\mathbf{g}}_0^1/N$$
$$= ts_0\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0 - ts_0\mathbf{E}\ \mathbf{g}^T H_0\mathbf{x}_1 y_1/N + t^2 s_0\mathbf{E}\ \mathbf{g}^T H_0\mathbf{x}_1\mathbf{x}_1^T H_0^1\widehat{\mathbf{g}}_0^1/N.$$

Here the first summand is that included into the lemma formulation. The square of the sum of the second and third terms on the right hand side is not greater than

$$\begin{aligned} &2t^2\mathbf{E}\ (\mathbf{g}^T H_0\mathbf{x}_1)^2[\mathbf{E}\ y_1^2 + t^2\mathbf{E}\ (\mathbf{x}_1^T H_0^1\widehat{\mathbf{g}}_0^1)^2]/N^2 \\ &\quad \le 2t^2\mathbf{g}^2\mathbf{E}\ u_1^2\sqrt{M}(1 + t^2\mathbf{E}\ (\widehat{\mathbf{g}}_0^1)^2)/N^2 \\ &\quad \le 2M\tau^2(1 + 2\tau^2(1+\lambda))/N^2 \le Mc_{41}/N^2. \end{aligned}$$

We conclude that $o_1^2 \le Mc_{63}\varepsilon^2$ and $|o_1| \le \sqrt{M}c_{32}\varepsilon$. In view of the first statement of Lemma 8.3, we have

$$t\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}} = ts_0\mathbf{E}\ \mathbf{g}^T H_0\mathbf{g} + o,$$

where $|o| \le |\mathbf{g}|c_{31}\varepsilon$. Consequently for $\sigma_2$ in (11) we have $|o_2| \le \sqrt{M}c_{32}\varepsilon$.

Now we estimate the variance of $f \stackrel{\text{def}}{=} t\widehat{\mathbf{g}}_0 H_0\widehat{\mathbf{g}}_0$. Using Lemma 2.2 and taking into account the identical dependence on sample vectors, we have var $f \le N\Delta_1^2$, where $\Delta_1 = f - f^1$, and $f^1$ does not depend on $\mathbf{x}_1$ and $y_1$. We rewrite $f$ in the form

$$\begin{aligned} f &= t\widehat{\mathbf{g}}_0^{1T} H_0^1\widehat{\mathbf{g}}_0^1 + 2t\widehat{\mathbf{g}}_0^{1T} H_0\mathbf{x}_1 y_1/N \\ &\quad + t\mathbf{x}_1^T H_0\mathbf{x}_1 y_1^2/N^2 - t^2\widehat{\mathbf{g}}_0^T H_0^1\mathbf{x}_1\mathbf{x}_1^T H_0\widehat{\mathbf{g}}_0/N. \end{aligned}$$

Here the first summand does not depend on $(\mathbf{x}_1, y_1)$. We find that var $f$ is not greater than

$$a\ [t^2\mathbf{E}\ (\widehat{\mathbf{g}}_0^{1T} H_0\mathbf{x}_1 y_1)^2 + t^2\mathbf{E}\ \psi_1^2 y_1^4 + t^4\mathbf{E}\ (\widehat{\mathbf{g}}_0^{1T} H_0^1\mathbf{x}_1)^2(\mathbf{x}_1^T H_0\widehat{\mathbf{g}}_0^1)^2]/N,$$

where $a$ is a numerical constant. We apply the Schwarz inequality. The square of the first summand in the square bracket is not greater than

$$Mt^4\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4\mathbf{E}\ y_1^4 \le 2M^4 t^4(1+\lambda)^2 \le M^2 c_{42}.$$

It follows that the first summand is not greater $Mc_{21}$ in absolute value. The second summand is not greater $Mt^2$ by (1). Using (6) and (1), we obtain that the third summand is not greater than

$$t^4\mathbf{E}\ (\widehat{\mathbf{g}}_0^T H_0^1 \mathbf{x}_1)^4/N^2 \leq Mt^4 \mathbf{E}\ |\widehat{\mathbf{g}}_0|^4 \leq 2M^3 t^4 (1+\lambda)^2 \leq Mc_{42}.$$

Consequently var $f \leq Mc_{42}/N$. This completes the proof of Lemma 8.4. □

LEMMA 8.5. *Let $A$ be a symmetric positive semidefinite matrix of constants. If $t \geq u \geq 0$, then*

$$tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t) A H_0(u) \widehat{\mathbf{g}}_0 = tu\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T} H_0^1(t) A H_0^1 \widehat{\mathbf{g}}_0^1 + o,$$

*where* $|o| \leq \sqrt{M} c_{31}/N$; *the inequality holds*

$$t^2 u^2 \text{var}\ (\widehat{\mathbf{g}}_0^T H_0(t) A H_0(u) \widehat{\mathbf{g}}_0) \leq M^2 t^2 c_{42}/N.$$

Proof. Substituting $\widehat{\mathbf{g}}_0 = \widehat{\mathbf{g}}_0^1 + \mathbf{x}_1 y_1/N$, we obtain

$$\begin{aligned} f \stackrel{\text{def}}{=} \widehat{\mathbf{g}}_0^T H_0(t) A H_0(u) \widehat{\mathbf{g}}_0 &= tu \widehat{\mathbf{g}}_0^{1T} H_0(t) A H_0(u) \widehat{\mathbf{g}}_0^1 \\ &+ tu\ \widehat{\mathbf{g}}_0^{1T} H_0(t) A H_0(u) \mathbf{x}_1 y_1/N + tu\ y_1 \mathbf{x}_1^T H_0(t) A H_0(u) \widehat{\mathbf{g}}_0^1/N \\ &+ tu\ y_1^2 \mathbf{x}_1^T H_0(t) A H_0(u) \mathbf{x}_1/N^2. \end{aligned} \tag{12}$$

To prove the lemma statement, it suffices to show that the three last summands in (12) are small and the difference

$$d = tu\ \widehat{\mathbf{g}}_0^{1T} H_0^1(t) A H_0^1(u) \widehat{\mathbf{g}}_0^1 - tu\ \widehat{\mathbf{g}}_0^{1T} H_0(t) A H_0(u) \widehat{\mathbf{g}}_0^1$$

is small. Using (4) we transform $d$ as follows:

$$\begin{aligned} d =& t^2 u\ \widehat{\mathbf{g}}_0^{1T} H_0^1(t) \mathbf{x}_1 \mathbf{x}_1^T H_0(t) A H_0^1(u) \widehat{\mathbf{g}}_0^1/N \\ &+ tu^2\ \widehat{\mathbf{g}}_0^{1T} H_0^1(u) \mathbf{x}_1 \mathbf{x}_1^T H_0(u) A H_0^1(t) \widehat{\mathbf{g}}_0^1/N \\ &+ t^2 u^2\ \widehat{\mathbf{g}}_0^{1T} H_0^1(t) \mathbf{x}_1 \mathbf{x}_1^T H_0(t) A H_0(u) \mathbf{x}_1 \mathbf{x}_1^T H_0^1 \widehat{\mathbf{g}}_0^1/N^2. \end{aligned}$$

Let us estimate $\mathbf{E}\ d^2$. Note that

$$\sqrt{tu}\mathbf{x}_1^T H_0(t) A H_0(u) \mathbf{x}_1 \leq \sqrt{tu\Phi(t)\Phi(u)} \leq 1.$$

It follows that

$$\begin{aligned}\mathbf{E}\ d^2/3 &\le t^4u^2\mathbf{E}\ |\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1|^2\mathbf{E}\ |\mathbf{x}_1^T H_0(t)AH_0^1(u)\widehat{\mathbf{g}}_0^1|^2/N^2\\ &+ t^2u^4\mathbf{E}\ |\widehat{\mathbf{g}}_0^{1T}H_0^1(u)\mathbf{x}_1|^2\mathbf{E}\ |\mathbf{x}_1^T H_0(u)AH_0^1(t)\widehat{\mathbf{g}}_0^1|^2/N^2\\ &+ Mt^3u^3\mathbf{E}\ |\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1|^2|\widehat{\mathbf{g}}_0^{1T}H_0^1(u)\mathbf{x}_1|^2/N^2. \qquad (13)\end{aligned}$$

Substituting $H_0^1$ for $H_0$, using the relation $H_0\mathbf{x}_1 = (1 - t\psi_1)H_0^1\mathbf{x}_1$ and (1), we obtain

$$\begin{aligned}\mathbf{E}\ d^2/3 &\le 2M^2t^6(\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^2)^2/N^2 + M^2t^6\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4/N^2\\ &\le 2M^4t^6(1+\lambda)^2/N^2 \le M^2t^2c_{42}/N^2.\end{aligned}$$

It follows that $\mathbf{E}\ |d| \le \sqrt{M}c_{31}/N$.

In the second summand in (12), we use the Schwarz inequality. First, we single out the dependence on $y_1$. The expected square of the second summand is not greater than

$$\begin{aligned}&2t^2u^2\sqrt{M}\mathbf{E}\ |\widehat{\mathbf{g}}_0^{1T}H_0^1(t)AH_0(u)\mathbf{x}_1|^2/N^2\\ &\qquad + 2t^4u^2\sqrt{M}\mathbf{E}\ (\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1)^2(\mathbf{x}_1^T H_0(t)AH_0(u)\mathbf{x}_1)^2/N^4.\end{aligned}$$

We use (6) to replace $H_0$ by $H_0^1$. It follows that the expected square of the second summand that is not greater than

$$4t^4M^2\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^2/N^2 \le 4M^2t^2\tau^2(1+\lambda)/N^2 \le M^2t^2c_{21}/N^2.$$

Therefore the second summand in (12) yields a contribution to $f$ not greater $\sqrt{M}c_{21}\varepsilon$.

The same estimate holds for the third term.

The contribution of the fourth summand in (12) to $f$ is not greater than $\sqrt{M}t\mathbf{E}\ y_1^2/N \le \sqrt{M}c_{10}/N$.

Thus the right hand side of (12) presents the leading term in the lemma formulation with the accuracy up to $c_{31}/N$. The first statement of the lemma is proved.

Further, we estimate var $f$ similiarly using Lemma 2.2. We find that var $f \le N\mathbf{E}\ \Delta_1^2$, where $f - \Delta_1$, does not depend on $\mathbf{x}_1$ and $y_1$. The value $\Delta_1$ equals the sum of last three terms in (12) minus $d$. The expectation of the squares of the second and third terms in (12) is not greater $M^2t^2c_{21}/N^2$. The square of the fourth term in (12), by (1), contributes no more than $Mt^2\mathbf{E}\ y_1^4/N^2 \le M^2t^2/N^2$. We have the inequality $\mathbf{E}\ d^2 \le 2M^2t^2c_{42}/N^2$. It follows that $\mathbf{E}\ \Delta_1^2 \le M^2t^2c_{42}/N^2$. This inequality gives the last inequality in (11). The proof of Lemma 8.5 is complete. □

LEMMA 8.6. *If $t \geq u \geq 0$, then*

$$\begin{aligned}&tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t) S H_0(u)\widehat{\mathbf{g}}_0 = tus_0(t)s_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T} H_0^1(t)\Sigma H_0^1(u)\widehat{\mathbf{g}}_0^1\\ &+(1-s_0(u))ts_0(t)\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{g} + (1-s_0(t))us_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T}H_0^1(u)\mathbf{g}\\ &+\sigma^2(1-s_0(t))(1-s_0(u)) + o,\\ &\mathbf{E}\ |o| \leq \sqrt{M}c_{42}\varepsilon; \quad t^2u^2 \mathrm{var}\ (\widehat{\mathbf{g}}_0^T H_0(t) S H_0(u)\widehat{\mathbf{g}}_0) \leq Mc_{62}/N.\end{aligned} \tag{14}$$

Proof. We notice that

$$f \stackrel{\text{def}}{=} tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t) S H_0(u)\widehat{\mathbf{g}}_0 = tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)\mathbf{x}_1\mathbf{x}_1 H_0(u)\widehat{\mathbf{g}}_0.$$

Substituting $\widehat{\mathbf{g}}_0 = \widehat{\mathbf{g}}_0^1 + \mathbf{x}_1 y_1/N$, we find

$$\begin{aligned}f =& tu\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T} H_0(t)\mathbf{x}_1\mathbf{x}_1^T H_0(u)\widehat{\mathbf{g}}_0^1 + tu\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T} H_0(t)\psi_1(u)\mathbf{x}_1 y_1\\ &+ tu\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T} H_0(u)\psi_1(t)\mathbf{x}_1 y_1 + tu\mathbf{E}\ \psi_1(t)\psi_1(u)y_1^2/N^2.\end{aligned}$$

In the first summand of the right hand side, we substitute $H_0\mathbf{x}_1$ from (6) and $t\psi_1(t) = 1 - s_0(t) + \Delta_1$, $\mathbf{E}\ \Delta_1^2 \leq \delta$. We find that the first summand is $tu\mathbf{E}\ s_0(t)s_0(u)\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\Sigma H_0^1(u)\widehat{\mathbf{g}}_0^1 + o$, where the leading term is involved in the formulation of the lemma, and the remainder term is such that

$$\mathbf{E}\ o^2 \leq at^2u^2\mathbf{E}\ \ [\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1\mathbf{x}_1^T H_0^1(u)\widehat{\mathbf{g}}_0^1]^2\delta,$$

where $a$ is numerical coefficient. In view of (1),

$$\mathbf{E}\ o^2 \leq Mt^4\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^4 c_{42}\varepsilon^2 \leq Mc_{84}\varepsilon^2.$$

We transform the three last summands of $f$ substituting $t\psi_1(t) = 1 - s_0(t) + \Delta_1$, $\mathbf{E}\ \Delta_1^2 \leq \delta$. The sum of these terms is equal to

$$\begin{aligned}&(1-s_0(u))ts_0(t)\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1 y_1\\ &\quad + (1-s_0(t))us_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T}H_0^1(u)\mathbf{x}_1 y_1\\ &\quad + (1-s_0(t))(1-s_0(u))\mathbf{E}\ y_1^2 + o,\end{aligned} \tag{15}$$

where the remainder term $o$ is such that $\mathbf{E}\ o^2$ is not greater than

$$a\{u^2\mathbf{E}\ (\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1)^2y_1^2\delta + t^2\mathbf{E}\ (\widehat{\mathbf{g}}_0^{1T}H_0^1(u)\mathbf{x}_1)^2y_1^2\delta + \mathbf{E}\ y_1^4\delta\}$$
$$\leq aM(2\tau^2(1+\lambda)+1)\delta \leq Mc_{63}\varepsilon^2,$$

where $a$ is a numerical coefficient. The leading part of (15), as is readily seen, coincides with three terms in (14). The weakest upper estimate for the squares of the remainder terms is $Mc_{84}\varepsilon^2$. Consequently the first lemma statement holds with the remainder term $\sqrt{M}c_{42}\varepsilon$.

To estimate the variance in the second statement, we first substitute $tSH_0(t) = I - H_0(t)$. It follows

$$\text{var}\ f = \text{var}\ (u\widehat{\mathbf{g}}_0^TH_0(t)\widehat{\mathbf{g}}_0 - u\widehat{\mathbf{g}}_0^TH_0(t)H_0(u)\widehat{\mathbf{g}}_0).$$

Here the variance of the minuend is not greater than $Mc_{42}/N$ by Lemma 8.4. The variance of the subtrahend is not greater $Mc_{62}/N$ by Lemma 8.5. The last statement of Lemma 8.6 follows. □

THEOREM 8.1. *If* $t \geq u \geq 0$*, then*

$$tu\mathbf{E}\ \widehat{\mathbf{g}}_0^TH_0(t)SH_0(u)\widehat{\mathbf{g}}_0 =$$
$$= tus_0(t)s_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^TH_0(t)\Sigma H_0(u)\widehat{\mathbf{g}}_0 + (1-s_0(u))ts_0(t)\mathbf{E}\ \widehat{\mathbf{g}}_0^TH_0(t)\mathbf{g}$$
$$+ (1-s_0(t))us_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^TH_0(u)\mathbf{g} + \sigma^2(1-s_0(t))(1-s_0(u)) + o, \tag{16}$$

*where* $|o| \leq \sqrt{M}c_{42}\varepsilon$.

Proof. First, we apply Lemma 8.6. The left hand side can be transformed by (14) with the remainder term $\sqrt{M}c_{42}\varepsilon$. We obtain the first summand in (16). Now we compare the right hand sides of (14) and (16). By (6), the difference between the second summands does not exceed

$$t|\mathbf{E}\ \widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{g} - \mathbf{E}\ \widehat{\mathbf{g}}_0^TH_0(t)\mathbf{g}|$$
$$\leq t|\mathbf{E}\ t\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1\mathbf{x}_1^TH_0(t)\mathbf{g}/N| + t|\mathbf{E}\ y_1\mathbf{x}_1^TH_0(t)\mathbf{g}/N|$$
$$\leq \{|\mathbf{E}\ t\widehat{\mathbf{g}}_0^{1T}H_0^1(t)\mathbf{x}_1|^2\mathbf{E}\ |t\mathbf{x}_1^TH_0(t)\mathbf{g}|^2 + \mathbf{E}\ y_1^2\ \mathbf{E}\ |t\mathbf{x}_1^TH_0(t)\mathbf{g}|^2\}^{1/2}/N$$
$$\leq \left(\mathbf{E}\ |\widehat{\mathbf{g}}_0^1|^2M^2t^4 + M^2t^2\right)^{1/2}/N \leq \sqrt{M}c_{21}/N.$$

The difference between the third summands also does not exceed this quantity. The fourth summands coincide. We conclude that the equality in the formulation of the theorem holds with the inaccuracy at most $\sqrt{M}c_{42}\varepsilon$. Theorem 8.1 is proved. □

### Functionals Depending on Sample Covariance Matrices and Sample Covariance Vectors

To pass to $C$ $H = H(t)$ and $\widehat{\mathbf{g}}$, we use the identities $C = S - \bar{\mathbf{x}}\bar{\mathbf{x}}^T$, $\widehat{\mathbf{g}} = \widehat{\mathbf{g}}_0 - \bar{\mathbf{x}}\bar{y}$, and the identity $H = H_0 - tH_0\bar{\mathbf{x}}\bar{\mathbf{x}}^T H$.

**Remark 2.** $\mathbf{E}\ |\widehat{\mathbf{g}}|^4 \leq aM^2(1+\lambda)^2$, where $a$ is a numerical coefficient.

Indeed, we notice that $\widehat{\mathbf{g}} = \widehat{\mathbf{g}}_0 - \bar{\mathbf{x}}\bar{y}$ and $|\widehat{\mathbf{g}}|^4 \leq 8\mathbf{E}\ |\widehat{\mathbf{g}}_0|^4 + 8\mathbf{E}\ |\bar{\mathbf{x}}|^4\bar{y}^4$. In view of Remark 1 it suffices to estimate $\mathbf{E}\ (\bar{\mathbf{x}}^2)^2\bar{y}^4$. This quantity is a sum of $n^2N^8$ summands. Summing over components of $\mathbf{x}$, we obtain a sum of $N^4$ products involving factors $(\mathbf{x}_{m1}^T\mathbf{x}_{m2})(\mathbf{x}_{m3}^T\mathbf{x}_{m4})$, $m1, m2, m3, m4 = 1, \ldots, N$. We majorize them by values $(\mathbf{x}_{m1}^2 + \mathbf{x}_{m2}^2)(\mathbf{x}_{m3}^2 + \mathbf{x}_{m4}^2)/4$. Thus the products are obtained that can be majorized by the Schwarz inequality with the quantities $\mathbf{E}\ (\mathbf{x}^2)^2y^4 \leq n^2M^2$. The sum of these is not greater $aM^2\lambda^2/N^2$ by (1). We conclude that Remark 2 is justified.

LEMMA 8.7.

$$t\mathbf{E}\ \mathbf{g}^T H(t)\widehat{\mathbf{g}} = ts(t)\mathbf{E}\ \mathbf{g}^T H(t)\mathbf{g} + o_1,$$
$$t\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)\widehat{\mathbf{g}} = \sigma^2(1 - s(t)) + ts(t)\mathbf{E}\ \mathbf{g}^T H(t)\widehat{\mathbf{g}} + o_2, \qquad (17)$$

*where* $|o_1|$, $|o_2| \leq \sqrt{M}c_{43}\varepsilon$.

Proof. We have

$$t\mathbf{E}\ \mathbf{g}^T H\widehat{\mathbf{g}} = t\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0 + t^2\mathbf{E}\ \mathbf{g}^T H\bar{\mathbf{x}}\bar{\mathbf{x}}^T H\widehat{\mathbf{g}}_0 - t\mathbf{E}\ \mathbf{g}^T H_0\bar{\mathbf{x}}\bar{y}.$$

By Lemma 8.3 the first summand equals $ts_0\mathbf{E}\ \mathbf{g}^T H_0\mathbf{g} + o$, where $|o| \leq \sqrt{M}c_{31}\varepsilon$. We estimate the remaining terms using the Schwarz inequality. By statement 3 of Lemma 8.1 with $\mathbf{e} = \mathbf{g}/|\mathbf{g}|$, we find that the second term is not greater in absolute value than

$$|\mathbf{g}|(\mathbf{E}\ tV^2t^3\mathbf{E}\ \bar{\mathbf{x}}^2\widehat{\mathbf{g}}^2)^{1/2} \leq \sqrt{M}c_{42}\varepsilon.$$

The third summand does not exceed $|\mathbf{g}|t(\mathbf{E}\ \bar{\mathbf{x}}^2\mathbf{E}\ \bar{y}^2)^{1/2} \le \sqrt{M}c_{11}$. We conclude that

$$|t\mathbf{E}\ \mathbf{g}^T H\widehat{\mathbf{g}} - t\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0| \le \sqrt{M}c_{42}\varepsilon.$$

In view of Lemma 8.1 we can replace $s_0$ by $s$ with an accuracy up to $t\mathbf{g}^2 c_{11}/N \le \sqrt{M}c_{21}/N$. It follows that $t\mathbf{E}\ \mathbf{g}^T H\widehat{\mathbf{g}} = ts\mathbf{g}^T H_0\mathbf{g} + o$, where $|o| \le \sqrt{M}c_{42}\varepsilon$. In view of statement 5 of Lemma 8.1, replacing $H_0$ by $H$ in the right hand side, we produce an inaccuracy of the same order. The first statement of our lemma is proved.

Further, from (16) if follows

$$t\mathbf{E}\ \widehat{\mathbf{g}}^T H\widehat{\mathbf{g}} = t\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0\widehat{\mathbf{g}}_0 + t^2\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0\bar{\mathbf{x}}\bar{\mathbf{x}}^T H\widehat{\mathbf{g}}_0 - 2t\mathbf{E}\ \bar{y}\bar{\mathbf{x}}^T H\widehat{\mathbf{g}},$$

where by Lemma 8.2 the second summand is not greater in absolute value than

$$t^2\left(\mathbf{E}\ |\widehat{\mathbf{g}}_0^T H_0\bar{\mathbf{x}}|^2\mathbf{E}\ \bar{\mathbf{x}}^2\widehat{\mathbf{g}}_0^2\right)^{1/2} \le \sqrt{M}c_{43}\varepsilon.$$

The third summand in absolute value is not greater than

$$t\ (\mathbf{E}\ \bar{y}^2\mathbf{E}\ \bar{\mathbf{x}}^2\widehat{\mathbf{g}}^2)^{1/2} \le \sqrt{M}c_{11}/\sqrt{N}.$$

Applying Lemma 8.4 we recall that

$$t\mathbf{E}\ \widehat{\mathbf{g}}_0^T H\widehat{\mathbf{g}}_0 = \sigma^2(1-s_0) + ts_0^2\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0 + o,$$

where $\mathbf{E}\ |o| \le \sqrt{M}c_{32}\varepsilon$. The difference between $s$ and $s_0$ contributes no more than $\sqrt{M}c_{21}/N$. Now we have

$$t\mathbf{E}\ \mathbf{g}^T H_0\widehat{\mathbf{g}}_0 = t\mathbf{E}\ \mathbf{g}^T H\widehat{\mathbf{g}} - t^2\mathbf{E}\ \mathbf{g}^T H_0\bar{\mathbf{x}}\bar{\mathbf{x}}^T H\widehat{\mathbf{g}}_0 + t\mathbf{E}\ \mathbf{g}^T H\bar{\mathbf{x}}\bar{y},$$

where the first summand is written out in the lemma formulation. By Lemma 8.2 with $\mathbf{e} = \mathbf{g}/|\mathbf{g}|$, the second term is not greater in absolute value than

$$|\mathbf{g}|\left(\mathbf{E}\ tV^2\mathbf{E}\ t^3\bar{\mathbf{x}}^2\widehat{\mathbf{g}}_0^2\right)^{1/2} \le \sqrt{M}c_{42}\varepsilon.$$

The third term in the absolute value is not greater than

$$t(\mathbf{E}\ \bar{\mathbf{x}}^2\mathbf{E}\ \bar{y}^2)^{1/2} \le \sqrt{M}c_{11}/N.$$

We conclude that the first part of statement 2 is valid. The second equation in statement 2 follows from statement 1, Lemma 8.1, and Lemma 8.4. This proves our Lemma 8.7. □

Lemma 8.8. *If $t \geq u \geq 0$, then*

$$tu|\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)\Sigma H(u)\widehat{\mathbf{g}} - \mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)\Sigma H_0(u)\widehat{\mathbf{g}}_0| \leq \sqrt{M} c_{63}\varepsilon.$$

Proof. Replacing $H$ by $H_0 - tH_0\bar{\mathbf{x}}\bar{\mathbf{x}}^T H$, we obtain

$$\begin{aligned} &tu\widehat{\mathbf{g}}^T H(t)\Sigma H(u)\widehat{\mathbf{g}} = \\ &\quad = tu\widehat{\mathbf{g}}^T H_0(t)\Sigma H_0(u)\widehat{\mathbf{g}} + t^2u\widehat{\mathbf{g}}^T H_0(t)\bar{\mathbf{x}}\bar{\mathbf{x}}^T H(t)\Sigma H_0(u)\widehat{\mathbf{g}} \\ &\quad + tu^2\widehat{\mathbf{g}}H_0(u)\bar{\mathbf{x}}\bar{\mathbf{x}}^T H(u)\Sigma H_0(t)\widehat{\mathbf{g}} \\ &\quad + t^2u^2\widehat{\mathbf{g}}^T H_0(t)\bar{\mathbf{x}}\bar{\mathbf{x}}^T H(t)\Sigma H(u)\bar{\mathbf{x}}\bar{\mathbf{x}}^T H_0(u)\widehat{\mathbf{g}}, \qquad (18) \end{aligned}$$

Here the first summand provides the required expression in the formulation of the lemma with an accuracy to the replacement of vectors $\widehat{\mathbf{g}}$ by $\widehat{\mathbf{g}}_0$, i.e., to an accuracy up to

$$tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)\Sigma H_0(u)\bar{\mathbf{x}}\bar{y} + tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(u)\Sigma H_0(t)\bar{\mathbf{x}}\bar{y} \\ + tu\mathbf{E}\ \bar{\mathbf{x}}^T H_0(t)\Sigma H_0(u)\bar{\mathbf{x}}\bar{y}^2.$$

Let us obtain upper estimates for these three terms. We single out first the dependence on $\bar{y}$. The square of the first term does not exceed

$$Mt^4\mathbf{E}\ \bar{y}^2\mathbf{E}\ (\widehat{\mathbf{g}}^2\bar{\mathbf{x}}^2) \leq 2M^3t^4\lambda(1+\lambda)/N \leq Mc_{42}/N.$$

The square of the second term can be estimated likewise. The square of the third term is not large than

$$Mt^2u^2\mathbf{E}\ \Phi(t)\Phi(u)\bar{y}^2 \leq Mt^2\mathbf{E}\ \bar{y}^2 \leq \sqrt{M}\tau^2/N.$$

It remains to estimate the sum of three last terms in (18). We have $\|\Sigma\| \leq M$. By Lemma 8.2 and Remark 1, for $u \leq t$, the expectation of the second summand in (18) is not greater than

$$\sqrt{M}\left(\mathbf{E}\ |t\widehat{\mathbf{g}}_0^T H_0\bar{\mathbf{x}}|^2 t^4\mathbf{E}\ \bar{\mathbf{x}}^2\widehat{\mathbf{g}}^2\right)^{1/2} \leq c_{53}\varepsilon.$$

The third summand can be estimated likewise. To estimate the fourth summand in (18), we note that

$$|\sqrt{u}H_0(u)\bar{\mathbf{x}}|^2 \leq |u\bar{\mathbf{x}}^T H_0^2(u)\bar{\mathbf{x}}| \leq |u\Phi(u)| \leq 1.$$

In view of Lemma 8.2, the expectation of the fourth summand in (18) is not larger

$$\sqrt{M}\left(\mathbf{E}\ |t\widehat{\mathbf{g}}^T H_0(t)\bar{\mathbf{x}}|^2\mathbf{E}\ t^5|\bar{\mathbf{x}}|^4\widehat{\mathbf{g}}^2\right)^{1/2} \leq \sqrt{M}c_{63}\varepsilon.$$

This is the weakest estimate. We conclude that the statement of Lemma 8.8 holds.

LEMMA 8.9. *If $t \geq u \geq 0$, then*

$$tu|\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)CH(u)\widehat{\mathbf{g}} - \mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)SH_0(u)\widehat{\mathbf{g}}_0| \leq \sqrt{M}c_{43}\varepsilon.$$

Proof. We substitute $\widehat{\mathbf{g}} = \widehat{\mathbf{g}}_0 - \bar{\mathbf{x}}\bar{y}$. It follows that

$$\begin{aligned} f \stackrel{\text{def}}{=} tu\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)CH(u)\widehat{\mathbf{g}} = \\ = tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H(t)CH(u)\widehat{\mathbf{g}}_0 - tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H(t)CH(u)\bar{\mathbf{x}}\bar{y} \\ - tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H(u)CH(t)\bar{\mathbf{x}}\bar{y} + tu\mathbf{E}\ \bar{y}^2\bar{\mathbf{x}}^T H_0(t)CH_0(u)\bar{\mathbf{x}}. \end{aligned} \tag{19}$$

Substituting $uCH(u) = I - H(u)$ in the last three last summands, we find that the square of the second term is not greater

$$t^2\mathbf{E}\ \bar{y}^2\mathbf{E}\ \widehat{\mathbf{g}}_0^2\bar{\mathbf{x}}^2 \leq 2M^2t^2\lambda(1+\lambda)/N \leq Mc_{22}/N.$$

The square of the third summand can be estimated likewise. The square of the fourth summand does not exceed

$$t^2\mathbf{E}\ \bar{y}^4\mathbf{E}\ |\bar{\mathbf{x}}|^4 \leq M^2t^2\lambda^2/N^2 \leq Mc_{22}/N^2.$$

Thus the quantity $f$ is equal to the first summand of the right hand side of (19) to an accuracy up to $\sqrt{M}c_{11}/\sqrt{N}$.

It remains to estimate the contribution of the difference

$$\begin{aligned} uH_0(t)CH(u) - uH_0(t)SH_0(u) \\ = H(t) - H(t)H(u) - H_0(t) + H_0(t)H_0(u). \end{aligned}$$

Using (16) we find that within an accuracy to $\sqrt{M}c_{11}/\sqrt{N}$,

$$\begin{aligned} |f - tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)SH_0(u)\widehat{\mathbf{g}}_0| \\ \leq t\mathbf{E}\ |\widehat{\mathbf{g}}_0^T(H(t) - H_0(t))(H(u) + H_0(u))\widehat{\mathbf{g}}_0| \\ = t^2\mathbf{E}\ |\widehat{\mathbf{g}}_0^T H_0(t)\bar{\mathbf{x}}\bar{\mathbf{x}}^T G\widehat{\mathbf{g}}_0| \leq 2\left(\mathbf{E}\ |t\widehat{\mathbf{g}}_0^T H_0(t)\bar{\mathbf{x}}|^2\mathbf{E}\ t^2\bar{\mathbf{x}}^2\widehat{\mathbf{g}}_0^2\right)^{1/2}, \end{aligned}$$

where $\|G\| \leq 2$. By Lemma 8.2 the right hand side of the last inequality is not larger $2Mt\sqrt{\lambda(1+\lambda)}c_{32}\varepsilon \leq \sqrt{M}c_{43}\varepsilon$. We conclude that the statement of our lemma is to an accuracy up to $\sqrt{M}c_{43}\varepsilon$. The lemma is proved. □

THEOREM 8.2. *If $t \geq u \geq 0$, then*

$$tu\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)CH(u)\widehat{\mathbf{g}} = tus(t)s(u)\mathbf{E}\ \widehat{\mathbf{g}}^T H(t)\Sigma H(u)\widehat{\mathbf{g}} + (1-s(u))ts^2(t)\mathbf{E}\ \mathbf{g}^T H(t)\mathbf{g} + (1-s(t))us^2(u)\mathbf{E}\ \mathbf{g}^T H(u)\mathbf{g} + \sigma^2(1-s(t))(1-s(u)) + o, \tag{20}$$

*where* $|o| \leq \sqrt{M}c_{63}\varepsilon$.

Proof. We transform the left hand side. First, we apply Lemma 8.9 and obtain the leading term $tu\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)SH_0(u)\widehat{\mathbf{g}}_0$ with a correction not greater than $\sqrt{M}c_{43}\varepsilon$. Then we apply Theorem 8.1. To the same accuracy it follows that

$$tus_0(t)s_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)\Sigma H_0(u)\widehat{\mathbf{g}}_0 + (1-s_0(u))ts_0(t)\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(t)\mathbf{g} + (1-s_0(t))us_0(u)\mathbf{E}\ \widehat{\mathbf{g}}_0^T H_0(u)\mathbf{g} + \sigma^2(1-s_0(t))(1-s_0(u)). \tag{21}$$

We transform the first summand in (21) using Lemma 8.5. This lemma gives a correction not greater than $\sqrt{M}c_{31}/N$. The first summand in the right hand side of (20) is obtained with a correction not greater than $\sqrt{M}c_{63}\varepsilon$. Next we transform the second summand in (21). By Lemma 8.5 the equality holds $t\mathbf{E}\ \widehat{\mathbf{g}}^T H_0(t)\mathbf{g} = ts_0(t)\mathbf{E}\ \mathbf{g}^T H_0(t)\mathbf{g} + o$, where $o$ is not greater in absolute value than $Mtc_{31}\varepsilon \leq \sqrt{M}c_{41}\varepsilon$. The difference between $s_0$ and $s$ yields a lesser correction. We obtain the second summand of the right hand side of (20). Analogously we transform the similiar expression with the argument $u$. We obtain the third summand in (20). The substitution of $s$ for $s_0$ gives a correction in the last summand that is not larger $\sqrt{M}c_{11}/N$. We conclude that the right hand sides of (21) and (20) coincide to an accuracy up to $\sqrt{M}c_{63}\varepsilon$. This proves Theorem 8.2. □

### The Leading Part of the Quadratic Risk and its Estimator

We first express the leading part of the quadratic risk $R^1$ in terms of sample characteristics, that is, as a function of $\mathbf{C}$ and $\widehat{\mathbf{g}}$. Our problem is to construct reliable estimators for the functions t$\mathbf{E}\ \widehat{\mathbf{k}}^T\mathbf{g}$ and $D(t,u) = tu\mathbf{E}\ \widehat{\mathbf{k}}^T\Sigma\widehat{\mathbf{k}}$ that are involved in the expression (3) for the quadratic risk. We consider the statistics

$$\widehat{s}(t) = 1 - nN^{-1} + N^{-1}\text{tr}\ (I+tC)^{-1}, \qquad \widehat{\kappa}(t) = t\widehat{\mathbf{g}}^T H(t)\widehat{\mathbf{g}},$$

$$\widehat{K}(t,u) \stackrel{\text{def}}{=} tu\widehat{\mathbf{g}}^T H(t)CH(u)\widehat{\mathbf{g}} = \frac{t\widehat{\kappa}(u) - u\widehat{\kappa}(t)}{t-u},$$
$$\widehat{\Delta}(t,u) = \widehat{K}(t,u) - \big(1-\widehat{s}(t)\big)\widehat{\kappa}(u) - \big(1-\widehat{s}(u)\big)\widehat{\kappa}(t) + \widehat{\sigma}^2\big(1-\widehat{s}(t)\big)\big(1-\widehat{s}(u)\big),$$

where $\widehat{K}(t,u)$ is extended by continuity to $t=u$.

**Remark 3.** If $t \geq u \geq 0$, then

$$s(t)s(u)\mathbf{E}\ D(t,u) = \mathbf{E}\ \widehat{\Delta}(t,u) + o, \quad \text{where} \quad \mathbf{E}\ |o| \leq \sqrt{M}\, c_{63}\varepsilon.$$

It is convenient to replace the dependence of the functionals on $\eta(t)$ by that on a function $\rho(t)$ of the form $\rho(t) \stackrel{\text{def}}{=} \int_{0\leq x\leq t} \frac{1}{s(x)}\, d\eta(x)$. We note that the variation of the function $t^k\rho(t)$ on $[0,\infty)$ does not exceed $\sqrt{M}\,\eta_{k+1}\lambda$. Let us consider the quadratic risk (2) as a function of $\rho(t)$, $R = \mathbf{E}\ \Delta^2 = R(\eta) = R(\rho)$.

THEOREM 8.3. *The statistic*

$$\widehat{R}(\rho) \stackrel{\text{def}}{=} \widehat{\sigma}^2 - 2\int \big[\widehat{\kappa}(t) - \widehat{\sigma}^2\big(1-\widehat{s}(t)\big)\big]\, d\rho(t) + \iint \widehat{\Delta}(t,u)\, d\rho(t)\, d\rho(u)$$

*is an estimator of $R = R(\rho)$ for which $\mathbf{E}\ \widehat{R}(\rho) = R(\rho) + o$, where $|o| \leq \sqrt{M}\,\eta_8 c_{05}\varepsilon$.*

Proof. The expression (3) equals $R(\rho)$ with an accuracy up to $\sqrt{M}/N$. Let us compare (3) with the right hand side of the expression for $R(\widehat{\rho})$ in the formulation of our theorem. We have $|\sigma^2 - \mathbf{E}\ \widehat{\sigma}^2| \leq 2\sqrt{M/N}$, $s(t) \geq (1+\tau\lambda)^{-1}$. By Lemma 8.3,

$$|\mathbf{E}\ t\mathbf{g}^T H(t)\widehat{\mathbf{g}} - \mathbf{E}\ (\widehat{\kappa}(t) - \sigma^2(1-s(t))/s(t)| \leq \sqrt{M}c_{54}\varepsilon.$$

The differential $d\rho(t) = s^{-1}(t)d\eta(t)$, and the variation of $\eta(t)$ is not larger than 1. One can see that the second summand in (3) equals the second term of the expression for $\widehat{R}(\rho)$ with an accuracy up to $\sqrt{M}c_{54}\varepsilon$. The third summand in (3), by Theorem 8.2, is equal to the third term of $\widehat{R}(\rho)$ with an accuracy to $\sqrt{M}c_{85}\varepsilon$. The coefficient $c_{85}$ increases not faster than $t^8$ with $t$. We come to the theorem's statement. □

Now we pass to the calculation of the non-random leading part of the quadratic risk.

Define

$$\phi(t) = t\widehat{\mathbf{g}}^T(I + t\Sigma)^{-1}\mathbf{g}, \quad \kappa(t) = \sigma^2(1 - s(t))^2 + s^2(t)\phi(ts(t)),$$
$$K(t,u) = \frac{t\kappa(u) - u\kappa(t)}{t - u},$$
$$\Delta(t,u) = K(t,u) - (1 - s(t))\kappa(u) - (1 - s(u))\kappa(t) + \sigma^2(1 - s(t))(1 - s(u)),$$

where the function $K(t,u)$ is extended by continuity to $t = u$.

In view of Lemma 8.4, the following statement holds.

**Remark 4.** $\mathbf{E}\ \widehat{\kappa}(t) = \kappa(t) + o$, where $|o| \le \sqrt{M}c_{43}\varepsilon$.

LEMMA 8.10. *If* $t \ge u \ge 0$, *then*

$$|\mathbf{E}\ \widehat{K}(t,u) - K(t,u)| \le c_{43}\sqrt{M\varepsilon},$$
$$|\mathbf{E}\ \widehat{\Delta}(t,u) - \Delta(t,u)| \le c_{34}\sqrt{M\varepsilon}.$$

Proof. First, let $|t - u| \ge \alpha > 0$. For these arguments, by Remark 4 we have $|\mathbf{E}\ \widehat{K}(t,u) - K(t,u)| \le \tau c_{43}/\alpha$. Let $|t - u| \le \alpha$. We expand the functions $\kappa(u)$ and $\widehat{\kappa}(u)$ to the Taylor series up to the second derivatives

$$\begin{aligned} &\mathbf{E}\ \widehat{K}(t,u) - K(t,u) = \\ &\quad = \Delta^{-1}\mathbf{E}\ [u\widehat{\kappa}(u) + u\Delta\widehat{\kappa}'(u) + u\Delta\widehat{\kappa}''(\xi) + \Delta\widehat{\kappa}(u)] \\ &\qquad - \Delta^{-1}[u\kappa(u) + u\Delta\kappa'(u) + u\Delta^2\kappa''(\zeta) + \Delta\kappa(u)], \end{aligned} \tag{22}$$

where $\xi$ and $\zeta$ are intermediate values of the arguments, $u \le \xi,\ \zeta \le 1$. Here

$$\begin{aligned} |\mathbf{E}\ \widehat{k}''(\xi)| &\le 2\mathbf{E}\ \widehat{\mathbf{g}}^T HCH\widehat{\mathbf{g}} + t|\mathbf{E}\ \widehat{\mathbf{g}}^T HCHCH\widehat{\mathbf{g}}| \\ &\le 3\mathbf{E}\ \widehat{\mathbf{g}}^T C\widehat{\mathbf{g}} \le aM^{3/2}\lambda(1 + \lambda), \end{aligned}$$

where $a$ is a numerical coefficient. We also find that

$$|s'(u)| \le N^{-1}\mathbf{E}\ \text{tr}\ (HCH) \le \sqrt{M}\lambda,$$
$$|s''(u)| \le N^{-1}\mathbf{E}\ \text{tr}\ (HCHCH) \le N^{-1}\mathbf{E}\ \text{tr}\ C^2 \le M\lambda,$$
$$|\phi'(us(u))| \le \mathbf{g}^T R\mathbf{g} + t(1+\tau\lambda)\mathbf{g}^T R\Sigma Rg \le Mc_{21},$$
$$|\phi''(us(u))| \le a(1+\tau\lambda)^2\mathbf{g}^T R\Sigma R\mathbf{g} \le M^{3/2}c_{22},$$

where $R = (I + us(u)\Sigma)^{-1}$. We conclude that $t|\mathbf{E}\ \kappa''(\xi)| \le Mc_{32}$. Thus the terms with the second derivatives contribute to (22) no more than $Mc_{32}\alpha$. Further, we estimate $|u\mathbf{E}\ \widehat{\kappa}'(u) - u\kappa'(u)|$. We have

$$\mathbf{E}\ \widehat{\kappa}(u+\Delta) - \mathbf{E}\ \widehat{\kappa}(u) = \Delta\widehat{\kappa}'(u) + \Delta^2\mathbf{E}\ \widehat{\kappa}''(\xi).$$

Analogously we substitute $\kappa(u+\Delta) - \kappa(u)$. Subtracting these expressions we find that

$$\Delta\mathbf{E}\ \widehat{\kappa}'(u) + \Delta^2\mathbf{E}\ \widehat{\kappa}''(\xi) - \Delta\mathbf{E}\ \kappa'(u) - \Delta^2\mathbf{E}\ \kappa''(\zeta)$$

is not greater $\sqrt{M}c_{43}\varepsilon$ in absolute value. Consequently,

$$|\mathbf{E}\ u\widehat{\kappa}'(u) - u\kappa'(u)| \le \sqrt{M}c_{43}\varepsilon + Mc_{32}\alpha.$$

It remains to estimate the summand $(\mathbf{E}\ u\widehat{\kappa}(u) - u\kappa(u))/\Delta$ in the right hand side of (22). By Lemma 8.7 this difference is not greater than $c_{53}\varepsilon/\Delta$ in absolute value, and consequently all the right hand side of (22) does not exceed $\sqrt{M}c_{43}\varepsilon + c_{32}(M\alpha + c_{22}\varepsilon/\alpha)$. Let us choose $\alpha = \sqrt{c_{22}\varepsilon/M}$. Then (since $\varepsilon \le 1$) the right hand side of (22) is not greater in absolute value than $\sqrt{M\varepsilon}c_{43}$. The first statement is proved.

Further, we have

$$\begin{aligned}|\mathbf{E}\ \widehat{\Delta}(t,u) - \Delta(t,u)| &\le |\mathbf{E}\ \widehat{K}(t,u) - K(t,u)| \\ &+ |r(t)\kappa(u) - \widehat{r}(t)\widehat{\kappa}(u) + r(u)\kappa(t) - \widehat{r}(u)\widehat{\kappa}(t)| \\ &+ |\sigma^2 r(t)r(u) - \widehat{\sigma}^2\widehat{r}(t)\widehat{r}(u)|,\end{aligned}$$

where $r(t) = 1 - s(t)$, $\widehat{r}(t) = 1 - \widehat{s}(t)$. The first summand is estimated in Lemma 8.10. Note that the variance of $n^{-1}\text{tr}\ H$ is not greater than $c_{20}/N$ and, therefore, $|\mathbf{E}\ \widehat{s}(t) - s(t)| \le c_{11}\varepsilon$. By Remark 4 the upper estimate $c_{43}\sqrt{M\varepsilon}$ also holds for $|\mathbf{E}\ \widehat{\Delta}(t,u) - \Delta(t,u)|$. Lemma 8.10 is proved. □

THEOREM 8.4. *The quadratic risk* (3) *is* $R = R(\rho) = R_0(\rho) + o$, *where*

$$R_0 = R_0(\rho) \stackrel{\text{def}}{=} \sigma^2 - 2\int s(t)\phi(ts(t))d\rho(t) + \iint \Delta(t,u)d\rho(t)d\rho(u), \tag{23}$$

*and* $|o| \le \sqrt{M}\eta_6 c_{05}\sqrt{\varepsilon}$. *If a function of bounded variation* $\rho^{\text{opt}}(t)$ *exists satisfying the equation*

$$\int \Delta(t,u)d\rho^{\text{opt}}(u) = \kappa(t) - \sigma^2(1 - s(t)), \quad t \ge 0,$$

*then* $R_0(\rho)$ *reaches the minimum for* $\rho(t) = \rho^{\text{opt}}(t)$ *and*

$$\min_\rho R_0(\rho) = \sigma^2 - \int s(t)\phi(ts(t))d\rho^{\text{opt}}(t).$$

Proof. We start from Theorem 8.3. The difference between $\sigma^2$ and $\mathbf{E}\,\widehat{\sigma}^2$ is not greater $2\sqrt{M/N}$. The difference between $s(t)$ and $\mathbf{E}\,\widehat{s}(t)$ does not exceed $c_{11}/\sqrt{N}$. By Lemma 8.7,

$$|\phi(ts(t)) - \mathbf{E}\,\left(\widehat{\kappa}(t) - \widehat{\sigma}^2(1 - \widehat{s}(t))\right)| \le \sqrt{M}c_{43}\varepsilon.$$

We obtain the two first summands in (23). Further, by Lemma 8.10, $|\Delta(t,u) - \mathbf{E}\,\widehat{\Delta}(t,u)| \le c_{43}\sqrt{M\varepsilon}$. The statement of Theorem 8.4 follows. □

Usually, to estimate the efficiency of the linear regression the residual sum of squares ('RSS') is used, which presents an empirical quadratic risk with $\mathbf{x}$ and $y$ from the same sample $\mathfrak{X}$

$$R^{\text{emp}} = \widehat{\sigma}^2 - 2\widehat{\mathbf{k}}^T\widehat{\mathbf{g}} + \widehat{\mathbf{k}}^T C\widehat{\mathbf{k}}.$$

THEOREM 8.5. *For the linear regression with* $\mathbf{k} = \Gamma\widehat{\mathbf{g}}$ *and* $l = \bar{y} - \widehat{\mathbf{k}}^T\bar{\mathbf{x}}$, *the empiric quadratic risk* $R^{\text{emp}} = R^{\text{emp}}(\eta)$ *can be written in the form*

$$R^{\text{emp}}(\eta) = \sigma^2 - 2\int \kappa(t)d\eta(t) + \iint K(t,u)d\eta(t)d\eta(u) + o,$$

*where* $|o| \le c_{43}\sqrt{M\varepsilon}$.

Proof. By Lemmas 8.3 and 8.10, we have

$$|\sigma^2 - \mathbf{E}\,\widehat{\sigma}^2| \le 2\sqrt{M}\varepsilon \quad |\kappa(t) - \mathbf{E}\,\widehat{\kappa}(t)| \le \sqrt{M}c_{43}\varepsilon,$$
$$|K(t,u) - \mathbf{E}\,\widehat{K}(t,u)| \le c_{43}\sqrt{M\varepsilon}.$$

The variation of $\eta(t)$ on $[0,\infty)$ is not larger 1. We can easily see that the statement of Theorem 8.5 holds. □

## Special cases

We consider 'shrinkage ridge estimators' defined by the function $\eta(x) = \alpha$ ind $(x \geq t)$, $x \geq 0$. The coefficient $\alpha$ is an analog of the 'shrinkage coefficient' in estimators of the Stein estimator type, and $1/t$ serves as a regularization parameter. In this case, by Theorem 8.4 the leading part of the quadratic risk (3) is

$$R_0(\rho) = R_0(\alpha, t) = \sigma^2 - 2\alpha\phi(ts(t)) + \alpha^2\Delta(t,t)/s^2(t).$$

If $\alpha = 1$, we have

$$R_0(\rho) = R_0(1, t) = \frac{1}{s^2(t)}\frac{d}{dt}[t(\sigma^2 - \kappa(t))].$$

In this case, the empirical risk is $R^{\mathrm{emp}}(t) = s^2(t)R_0(t)$. For the optimal value $\alpha = \alpha^{\mathrm{opt}} = s^2(t)\phi(ts(t))/\Delta(t,t)$, we have

$$R_0(\rho) = R_0(\alpha^{\mathrm{opt}}, t) = \sigma^2\left(1 - \frac{s^2(t)\phi^2(ts(t))}{\Delta(t,t)}\right).$$

**Example 1.** Let $\lambda \to 0$ (the transition to the case of fixed dimension under the increasing sample size $N \to \infty$). To simplify formulas we set $\lambda = 0$ and write out only leading terms of the expressions. Then $s(t) = 1$, $h(t) = n^{-1}\mathrm{tr}\ (I + t\Sigma)^{-1}$, $\kappa(t) = \phi(t)$, $\Delta(t,t) = \phi(t) - t\phi'(t)$. Set $\Sigma = I$. We have

$$\phi(t) = \frac{\sigma^2 r^2 t}{1+t}, \quad h(t) = \frac{1}{1+t}, \quad \Delta(t,t) = \frac{\sigma^2 r^2 t^2}{(1+t)^2},$$

where $r^2 = \mathbf{g}^2/\sigma^2$ is the square of the multiple correlation coefficient. The leading part of the quadratic risk (3) is $R_0 = \sigma^2[1 - 2\alpha^2 t/(1+t) + \alpha^2 r^2 t^2/(1+t)^2]$. For the optimal choice of $\alpha$, as well as for the optimal choice of $t$, we have $\alpha = (1+t)/t$ and $R^{\mathrm{opt}} = \sigma^2(1-r^2)$, i.e., the quadratic risk (3) asymptotically attains its a priori minimum.

**Example 2.** Let $N \to \infty$ and $n \to \infty$ so that $\lambda = n/N \to \lambda^*$. Assume that the matrices $\Sigma$ are non-degenerate for each $n$, $\sigma^2 \to \sigma_*^2$, $r^2 = \mathbf{g}^T\Sigma^{-1}\mathbf{g}/\sigma^2 \to r_*^2$, and the parameters $\gamma \to 0$. Under the limit transition, for each fixed $t \geq 0$, the remainder terms in Theorems 8.2-8.5 vanish. Let $\alpha = 1$ and $t \to \infty$ (the transition to the standard non-regularized regression under the increasing dimension asymptotics). Under these conditions,

$$s(t) \to 1 - \lambda_*, \quad s'(t) \to 0, \quad \phi(ts(t)) \to \sigma_*^2 r_*^2,$$

$$\kappa(t) \to \kappa(\infty) \stackrel{\text{def}}{=} \sigma_*^2 r_*^2 (1 - \lambda_*) + \sigma_*^2 \lambda_*, \quad t\kappa'(t) \to 0.$$

The quadratic risk (3) tends to $R_*$ so that

$$\lim_{t\to\infty} \lim_{\gamma\to 0} \lim_{N\to\infty} |\mathbf{E}\ R(t) - R_*| = 0,$$

where $R_* \stackrel{\text{def}}{=} \sigma_*^2(1-r_*^2)/(1-\lambda_*)$. This limit formula was obtained by I.S. Yenyukov (see in [16]). It presents an explicit dependence of the quality of the standard regression procedure on the dimension of observations and the sample size. Note that under the same conditions the empirical risk $R^{\text{emp}} \to \sigma_*^2(1 - r_*^2)(1 - \lambda_*)$.

**Example 3.** Under the same conditions as in Example 2, let the coefficients $\alpha$ be chosen optimally and then $t \to \infty$. We have $\alpha = \alpha^{\text{opt}}(t) = s^2(t)\phi(ts(t))/\Delta(t,t)$ and $t \to \infty$. Then

$$s(t) \to 1 - \lambda_*, \quad \phi(ts(t)) \to \sigma_*^2 r_*^2,$$
$$\Delta(t,t) \to \sigma_*^2(1 - \lambda_*)[\lambda_*(1 - r_*^2) + (1 - \lambda_*)r_*^2],$$
$$\alpha^{\text{opt}} \to r_*^2(1 - \lambda_*)[\lambda_*(1 - r_*^2) + (1 - \lambda_*)r_*^2].$$

By (23) the quadratic risk (3) $R_0(t, \alpha^{\text{opt}}) \to R_*$ as $t \to \infty$, where

$$R_* = \frac{\sigma_*^2(1 - r_*^2)[\lambda_* + (1 - \lambda_*)r_*^2]}{\lambda_*(1 - r_*^2) + (1 - \lambda_*)r_*^2} \leq \frac{\sigma_*^2(1 - r_*^2)}{1 - \lambda_*}.$$

If $\lambda_* = 1$ the optimal shrinkage coefficient $\alpha^{\text{opt}} \to 0$ in such a way that the quadratic risk remains finite (tends to $\sigma_*^2$) is spite of the absence of a regularization, whereas the quadratic risk for the standard linear regression tends to infinity.

**Example 4.** Let $n$ and $N$ be fixed and $\Sigma = I$. Then by Lemma 8.1 the functions $h(t)$ and $s(t)$ are connected by the equation $h(t)(1+ts(t)) = 1 + o$, where $|o| \leq c_{32}\varepsilon$. The functions

$$\phi(ts(t)) = \sigma^2 r^2 (1 - h(t)) + o_1,$$
$$\kappa(t) = \sigma^2[r^2 s(t) + \lambda](1 - h(t)) + o_2,$$

where $r^2 = \mathbf{g}^2/\sigma^2$ and $|o_\nu| \leq \sqrt{M}c_{43}\varepsilon$, $\nu = 1, 2$. Set $\alpha = 1$. Solving the equation $h(1 + ts) = 1$ for $h = h(t)$ and $s = s(t)$ we can find the derivative $h'(t)$ and single out the leading part of $t\kappa'(t)$. Let, for example, $\lambda = 1$. Then $\kappa'(t) = h^3(1 - r^2 + r^2 h)/(2 - h) + o(1)$. We find that the leading part of (3) is equal to

$$R = R(h) = \frac{\sigma^2(1 - r^2(1 - h))}{h(2 - h)}.$$

For $r^2 > 5/6$ this function has a minimum for $h = h^{\text{opt}} < 1$ and $R(h^{\text{opt}}) = \sigma^2(1 - r^2)/(h^{\text{opt}})^2$. If $r^2 < 5/6$, the function $R(h)$ is monotone and its maximum is reached at $h = 1$, $R(1) = \sigma^2$.