CHAPTER V

ESTIMATION OF HIGH-DIMENSIONAL INVERSE COVARIANCE MATRICES

In this Chapter we construct stable asymptotically unimprovable estimators for increasing inverse covariance matrices.

We consider a hypothetical sequence of problems

$$\mathfrak{P} = \{(\mathfrak{S},\ \Sigma,\ N,\ \mathfrak{X},\ C,\ \widehat{\Sigma}^{-1})_n\}, \quad n = 1, 2, \ldots, \tag{1}$$

where $\mathfrak{S}$ is a population with the covariance matrix $\Sigma = \text{cov}\ (\mathbf{x}, \mathbf{x})$, $\mathfrak{X}$ is a sample of size N from $\mathfrak{S}$, and $\widehat{\Sigma}^{-1}$ is an estimator of $\widehat{\Sigma}^{-1}$ calculated using the sample covariance matrix

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

where $\bar{\mathbf{x}}$ is the sample mean (we do not write out the subscripts n for the arguments of (1)).

We search for estimators minimizing the quadratic risk

$$\mathbf{E}\ n^{-1} \text{tr}\ (\Sigma^{-1} - \widehat{\Sigma}^{-1})^2$$

in some classes of estimators.

Note that the minimum quadratic risk estimators of Σ^{-1} may be of immediate interest in some standard multivariate problems. For example, for linear sample regression with random predictors, known covariance vectors $\mathbf{g}$, and estimators $\widehat{\Sigma}^{-1}$ of the inverse covariance matrix Σ^{-1}, the quadratic risk is

$$\text{const} + \mathbf{E}\ \mathbf{g}^T (\Sigma^{-1} - \widehat{\Sigma}^{-1}) \Sigma (\Sigma^{-1} - \widehat{\Sigma}^{-1}) \mathbf{g}.$$

Averaging over all directions we obtain that the second summand is not greater than

$$\lambda_{\max}\ (\Sigma)\ |\mathbf{g}|\ \mathbf{E}\ n^{-1} \text{tr}\ (\Sigma^{-1} - \widehat{\Sigma}^{-1})^2.$$

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Shrinkage Estimators of the Inverse Covariance Matrices

For the beginning, we consider a special problem of improving scalar multiple estimators of the inverse covariance matrices for normal populations.

Let $\mathfrak{K}^{(1)}$ be a class of impractical estimators of the form $\widehat{\Sigma}^{-1} = \alpha C^{-1}$, where α is a non-random scalar.

We assume that the sequence $\mathfrak{P}$ of problems is restricted by the following conditions.

1/. For each n the observation vectors $\mathbf{x} \sim \mathbf{N}(0, \Sigma)$ and all eigenvalues of Σ are located on a segment $[c_1, c_2]$ where $c_1 > 0$ and c_2 do not depend on n.

2/. The values $\Lambda_{-\nu} = n^{-1}\mathrm{tr}\ \Sigma^{-\nu},\ \nu = 1, 2, 3, 4,$ do not depend on n (for simplicity of notations).

3/. For each n in $\mathfrak{P}$, the inequality $N = N(n) > n + 2$ is valid, and the ratios $n/N \to y < 1$ as $n \to \infty$.

Remark 1. Under assumptions 1–3 the limits exist

$$M_{-1} = \operatorname*{l.i.m.}_{n\to\infty}\ n^{-1}\mathrm{tr}\ C^{-1} = (1-y)^{-1}\Lambda_{-1},$$
$$M_{-2} = \operatorname*{l.i.m.}_{n\to\infty}\ n^{-1}\mathrm{tr}\ C^{-2} = (1-y)^{-2}\Lambda_{-2} + y(1-y)^{-3}\Lambda_{-1}^2 \quad (2)$$

Here (and in the following) the limits in the mean denote limits in the square mean and we denote this convergence by $\stackrel{2}{\to}$.

For $\tilde{\Sigma}^{-1} \in \mathfrak{K}^{(1)}$ the quadratic risk is

$$R_n = R_n(\alpha) = \mathbf{E}\ n^{-1}\mathrm{tr}\ (\Sigma^{-1} - \alpha C^{-1})^2.$$

For the standard estimator, $\alpha = 1$, and

$$R_n \to y^2(1-y)^{-2}\Lambda_{-2} + y(1-y)^{-3}\Lambda_{-1}^2.$$

Let us find the parameter α minimizing

$$R = R(\alpha) = \operatorname*{l.i.m.}_{n\to\infty}\ n^{-1}\mathrm{tr}\ (\Sigma^{-1} - \alpha C^{-1})^2.$$

Remark 2. Under assumptions 1-3 the value $R(\alpha)$ is minimum for $\alpha = \alpha^{\mathrm{opt}} = (1-y)^{-1}\Lambda_{-2}/M_{-2} = 1 - y - yM_{-1}^2/M_{-2}$ and

$$R(\alpha^{\mathrm{opt}}) = \left(\frac{(1-y)^2\Lambda_{-2}}{\Lambda_{-2} + y(1-y)^{-1}\Lambda_{-1}^2}\right)\left(\frac{\Lambda_{-1}^2}{\Lambda_{-1}^2 + y(1-y)\Lambda_{-2}}\right) R(1).$$

However, the constant α^{opt} is unknown to the observer. Consider a class $\mathfrak{K}^{(2)}$ of estimators of the form $\widehat{\Sigma}^{-1} = \widehat{\alpha}_n\, C^{-1}$, where $\widehat{\alpha}_n$ tends to a constant in the square mean as $n \to \infty$.

Remark 3. Under assumptions 1-3 for the estimator $\tilde{\Sigma}^{-1} = \widehat{\alpha}_n C^{-1}$ with $\widehat{\alpha}_n \xrightarrow{2} \alpha$, we have $\lim\limits_{n\to\infty} R_n(\widehat{\alpha}_n) = R(\alpha)$.

To estimate α^{opt}, we construct the statistic

$$\widehat{\alpha}^{\mathrm{opt}} = \widehat{\alpha}^{\mathrm{opt}}(C) = \max\ \left(0, 1 - \frac{n}{N} - \frac{1}{N}\frac{\mathrm{tr}\ ^2 C^{-1}}{\mathrm{tr}\ C^{-2}}\right).$$

Remark 4. Under assumptions 1-3, the estimators $\widehat{\Sigma}^{-1} = \widehat{\alpha}_n^{\mathrm{opt}} C^{-1} \in \mathfrak{K}^{(1)}$ and are such that

$$\underset{n\to\infty}{\mathrm{l.i.m.}}\ \widehat{\alpha}^{\mathrm{opt}} \xrightarrow{2} \alpha^{\mathrm{opt}},$$

$$\lim_{n\to\infty} R_n(\widehat{\alpha}^{\mathrm{opt}}) = R(\alpha^{\mathrm{opt}}) = \inf_{\mathfrak{K}^{(2)}}\ \lim_{n\to\infty} R_n(\widehat{\alpha}_n).$$

In this sense, the estimators $\widehat{\alpha}_n^{\mathrm{opt}}$ of the matrices Σ^{-1} asymptotically dominate estimators from the class $\mathfrak{K}^{(2)}$.

Generalized Ridge Estimators of the Inverse Covariance Matrices

It is well known that the regularized ‘ridge’ estimators of Σ^{-1} of the form $(C + \alpha I)^{-1}$ are often used to stabilize the inversion of sample covariance matrix C, where I is the identity matrix and $\alpha > 0$ is a regularization parameter [1], [2]. Usually the value of α is to be chosen empirically. We set the problem of search for the best estimator of Σ^{-1} in the class $\mathfrak{K}^{(3)}$ of generalized ridge-estimators of the form $\widehat{\Sigma}^{-1} = \Gamma(C)$, where

$$\Gamma = \Gamma(C) = \int_{t\geq 0} (I + tC)^{-1} d\eta(t),$$

and $\eta(t)$ is a function of a finite variation on $[0, \infty)$.

We consider populations with four moments of all variables in the sequence of problems (1). Let $M > 0$ and γ be the parameters defined in Chapter 2 that determine the applicability of the spectral theory of large sample covariance matrices.

We restrict $\mathfrak{P}$ with the following requirements.

A. For each n, the parameters $M < c_0$ and all eigenvalues of Σ lie on a segment $[c_1, c_2]$, where $c_0,\ c_1$, and $c_2 < \sqrt{c_0}$ do not depend on n.

B. The parameters γ vanish as $n \to \infty$.

C. The ratio $n/N \to y$, where $0 < y < 1$.

D. For $u \geq 0$ the functions

$$F_{0n}(u) = n^{-1} \sum_{i=1}^{n} \text{ ind } (\lambda_i \leq u) \to F_0(u), \tag{3}$$

where λ_i are eigenvalues of $\Sigma,\ \ i = 1, \ldots, n$.

The assumptions A–D provide the applicability of theorems that were proved in Chapter 2. Under these assumptions if $c_1 > 0$, then the limits exists $\Lambda_k = \lim\limits_{n\to\infty} n^{-1}\text{tr } \Sigma^k$ for any k.

Define a bounded region $\mathfrak{G}$ of complex z

$$\mathfrak{G} = \mathfrak{G}(\beta, \delta) = \{z : (|z| \leq \beta)\&(\text{Re } z \leq 0 \text{ or } |\text{Im } z| > \delta > 0)\}.$$

We will use some necessary results stated in Chapter 2. Let us formulate these in the form of a lemma.

LEMMA 5.1. *Under assumptions A-D the following is valid.*

1/. *If $z \in \mathfrak{G}$, then function $h(z)$ exists such that*

$$h(z) = \underset{n\to\infty}{\text{l.i.m.}}\ n^{-1}\text{tr } (I - zC)^{-1} = \int (1 - zs(z)u)^{-1} dF_0(u), \tag{4}$$

where $s(z) = 1 + y(h(z) - 1)$.

2/. *If $z \in \mathfrak{G}$ then*

$$\mathbf{E}\ (I - zC)^{-1} = (I - zs(z)\Sigma)^{-1} + O(N^{-1})K_n,$$

where K_n are matrices of constants with spectral norms uniformly bounded in $\mathfrak{G}$.

3/. *If $z \in \mathfrak{G}$ then*

$$\max_{i,j}\ |\mathbf{E}\ H_{ii}H_{jj} - \mathbf{E}\ H_{ii}\mathbf{E}\ H_{jj}| = O(N^{-1})$$

uniformly in $\mathfrak{G}$, $i, j = 1, \ldots, n$, *as* $n \to \infty$, *where* H *with subscripts denote entries of the matrix* $H = (I - zC)^{-1}$.

4/. *If* $c_1 > 0$ *and* $y > 0$, *then the function* $h(z)$ *is regular for* $\operatorname{Im} z \neq 0$ *and satisfies the Hölder inequality*

$$|h(z) - h(z')| < c_3 |z - z'|^{\zeta},$$

where $\zeta > 0$ *is a numerical constant,* $0 < \zeta < 1$.

5/. *If* $c_1 > 0$, *then as* $|z| \to \infty$ *we have*

$$\begin{aligned} zh(z) &= -(1-y)^{-1}\Lambda_{-1} + O(|z|^{-1}) \text{ if } y < 1, \\ zh^2(z) &= -\Lambda_{-1} + O(|z|^{-1/2}) \text{ if } y = 1, \\ zs(z) &= -\lambda_0 + O(|z|^{-1}) \text{ if } y > 1, \end{aligned}$$

where β_0 *is a root of the equation*

$$\int (1 + \beta_o u)^{-1} dF_0(u) = 1 - y^{-1}.$$

6/. *As* $n \to \infty$, *the random functions*

$$F_n(u) = n^{-1} \sum_{j=1}^{n} \text{ind } (\lambda_j \leq u) \xrightarrow{2} F(u), \quad u \geq 0,$$

where λ_j *are eigenvalues of* C. *If* $c_1 > 0$ *and* $y > 0$, *then for each* $u > 0$ *there exists the derivative* $F'(u)$ *such that* $F'(u) \leq (c_1 y u)^{-1/2}$*; if* $y > 0$ *and* $0 < u < u_1$ *or* $u > u_2$, *where* $u_1 = c_1(1 - \sqrt{y})$ *and* $u_2 = c_2(1 + \sqrt{y})^2$, *then the derivative* $F'(u) = 0$.

7/. *The function* $h(z) = \int (1 - zu)^{-1} dF(u)$.

We introduce two additional assumptions:

$$1^o \; c_1 > 0 \quad \text{and} \quad 2^o \; \mathbf{E}\, n^{-1} \text{tr } C^{-6} \leq c_4, \quad n = 1, 2, \ldots \tag{5}$$

For normal populations the second condition follows from the fact that $c_1 > 0$.

Lemma 5.2. *Suppose conditions A–D and* (5) *hold in* $\mathfrak{P}$. *Then the limits exist*

$$M_{-1}= \underset{n\to\infty}{\text{l.i.m.}}\ n^{-1}\text{tr } C^{-1} = (1-y)^{-1}\Lambda_{-1},$$
$$M_{-2}= \underset{n\to\infty}{\text{l.i.m.}}\ n^{-1}\text{tr } C^{-2} = (1-y)^{-2}\Lambda_{-2} + y(1-y)^{-3}\Lambda_{-1}^2$$

Proof. We start from equation (4). For $z = it$ and a large $t > 0$ the function $h(z)$ is equal to

$$-(1-y)^{-1}\Lambda_{-1}z^{-1} - ((1-y)^{-2}\Lambda_{-2} + y(1-y)^{-3}\Lambda_{-1}^2)z^{-2} + \xi(z),$$
$$\text{where} \quad |\xi(z)| < c_1^{-3}t^{-3},$$
$$n^{-1}\text{tr } (I-zC)^{-1} = -n^{-1}\text{tr } C^{-1}z^{-1} - n^{-1}\text{tr } C^{-2}z^{-2} + \zeta(z),$$
$$\text{where } \ \mathbf{E}\ |\zeta(z)|^2 \le t^{-6}\mathbf{E}\ n^{-1}\text{tr } C^{-6} < c_4 t^{-6}.$$

By Lemma 5.1 we have $n^{-1}\text{tr } (I-zC)^{-1} \to h(z)$ as $n \to \infty$ in the square mean. Comparing these expressions as $t \to \infty$, we obtain the limits in the formulation of Lemma 5.2. □

We consider the statistics

$$\widehat{\Lambda}_{-1} \stackrel{\text{def}}{=} (n^{-1} - N^{-1})\ \text{tr } C^{-1},$$
$$\widehat{\Lambda}_{-2} \stackrel{\text{def}}{=} (1-nN^{-1})^2 n^{-1}\text{tr } C^{-2} + (1-nN^{-1})n^{-1}N^{-1}\text{tr }^2 C^{-1}.$$

Remark 5. Under conditions A–D and (5) as $n \to \infty$, we have $\widehat{\Lambda}_{-1} \stackrel{2}{\to} \Lambda_{-1}$, and $\widehat{\Lambda}_{-1} \stackrel{2}{\to} \Lambda_{-2}$

Denote

$$R_n = R_n(\Gamma) = \mathbf{E}\ n^{-1}\ \text{tr } (\Sigma^{-1} - \Gamma(C))^2. \tag{6}$$

Theorem 5.1. *Let conditions A–D and* (5) *hold. Then the limits exists* $\underset{n\to\infty}{\text{l.i.m.}}\ \widehat{\Lambda}_{-\nu} = \Lambda_{-\nu}$, $\nu = 1, 2$, *and*

$$R(\Gamma) \stackrel{\text{def}}{=} \lim_{n\to\infty} R_n(\Gamma) = \Lambda_{-2} - 2\int (\Lambda_{-1} - ts(-t)h(-t))d\eta(t) + \\ + \iint \frac{th(-t) - t'h(-t')}{t-t'}\ d\eta(t)d\eta(t'). \tag{7}$$

where the last integrand equals $d(th(-t))/dt$ *for* $t = t'$.

Proof. The first statement follows from Lemma 5.2. Next we have by definition

$$R_n(\Gamma) = n^{-1}\text{tr }\Sigma^{-2} - 2\mathbf{E}\ n^{-1}\text{tr }\Sigma^{-1}\Gamma(C) + \mathbf{E}\ n^{-1}\text{tr }\Gamma^2(C). \quad (8)$$

The first summand in the right hand side of (8) tends to Λ_{-2}. Let T be positive and large. By statement 2 of Lemma 5.1, the expectation of the second summand of the right hand side of (8) equals

$$-2n^{-1}\text{tr }[\Sigma^{-1}\int\limits_{t<T}\mathbf{E}\ Hd\eta(t)] + o_T =$$
$$= -\,2n^{-1}\text{tr }[\Sigma^{-1}\int\limits_{t<T}(I + ts\Sigma^{-1})d\eta(t)] + o_n(T) + o_T, \quad (9)$$

where $H = (I + tC)^{-1}$, $s = s(-t)$, $o_n(T) \to 0$ as $n \to \infty$ for any fixed $T > 0$. The value o_T is a contribution of $t \geq T$ that vanishes uniformly in n as a consequence of the decrease of the variation of $\eta(t)$ as $t \to \infty$. We use once more the expression for $\mathbf{E}\ H$ in Lemma 5.1 and find that (9) can be rewritten in the form

$$-2\int_{t<T} n^{-1}\text{tr }(\Sigma^{-1} - ts\mathbf{E}\ H)d\eta(t) + o'_n(T) + o_T =$$
$$= -2\int_{t<T}(\Lambda_{-1} - tsh)d\eta(t) + o''_n(T) + o_T, \quad (10)$$

where $h = h(-t)$, $s = s(-t)$, $o'_n(T) \to 0$ and $o''_n(T) \to 0$ as $n \to \infty$ for any fixed $T > 0$, and $o_T \to 0$ as $T \to \infty$. In view of statement 4 of Lemma 5.1, the quantity $|tsh|$ is bounded as $t \to \infty$, and as $n \to \infty$ and $T \to \infty$, we obtain the second term of the right hand side of (7).

Further, we consider

$$n^{-1}\text{tr }\Gamma^2(C) = \iint_{|t-t'|>\varepsilon} n^{-1}\text{tr }\frac{tH(-t) - t'H(-t')}{t' - t}d\eta(t)d\eta(t'), \quad (11)$$

where $H(-t) = (I + tC)^{-1}$, $\varepsilon > 0$. In the region where $|t - t'| > \varepsilon$, we have $h_n(-t) \stackrel{\text{def}}{=} n^{-1}\text{tr } H(-t) \stackrel{2}{\to} h(-t)$ as $n \to \infty$ for fixed $\varepsilon > 0$ and $t < T$ uniformly in t. Therefore the integrand in (11) converges to $(th(-t) - t'h(-t'))/(t - t')$ in the square mean uniformly, and we obtain the leading part of the last term in (7).

Now let us prove that the contribution of the region $|t - t'| < \varepsilon$ is small. We expand the integrand in (11) with respect to $x = |t - t'|$ near the point $t > 0$. Note that

$$|\frac{d}{dt}(th_n(-t))| \leq 1, \quad \mathbf{E}\ |\frac{d^2}{dt^2}th_n(-t)| \leq 2\mathbf{E}\ n^{-1}\text{tr } C^{-1} < c,$$

where, by (5), the quantity c does not depends on n and t. The expression in the left hand side of (11) is equal to

$$\frac{d}{dt}(th_n(-t)) + o_n(\varepsilon, T) + o_\varepsilon + o_T$$

where $o_n(\varepsilon, T) \to 0$ for fixed $\varepsilon > 0$ and $T > 0$ as $n \to \infty$; $o_\varepsilon \to 0$ as $\varepsilon \to +0$; and $o_T \to 0$ as $T \to \infty$ uniformly. We conclude that $\mathbf{E}\ n^{-1}\text{tr } \Gamma^2(C)$ converges to the third term of the right hand side of (7). This completes the proof of Theorem 5.1. $\square$

Denote $s_n(z) = 1 + nN^{-1}\,(h_n(z) - 1)$ and let $F_n(u)$ be the empiric distribution function of eigenvalues of C defined in statement 6 of Lemma 5.1.

Remark 6. Under assumptions A–D for $u > 0$ almost everywhere $F_n(u) \stackrel{2}{\to} F(u)$, and for any bounded continuous function $\varphi(u)$,

$$\underset{n\to\infty}{\text{l.i.m.}} \int \varphi(u)dF_n(u) = \int \varphi(u)dF(u).$$

The convergence of $F_n(u) \to F(u)$ in probability follows from Corollary 3.2.1 of Theorem 3.2.3 in [19] as a consequence of the convergence of $\{h_n(z)\}$ in probability. This also proves the weak convergence $\mathbf{E}F_n(u) \to F(u)$. The convergence of integrals with respect to non-random distribution is the property of the weak convergence. The convergence in the square mean follows from the boundedness. Remark 6 is justified. $\square$

Define

$$R(\Gamma) = \Lambda_{-2} - 2(1-y)\int u^{-1}\Gamma(u)dF(u) + 2y\iint \frac{\Gamma(u)-\Gamma(u')}{u'-u}dF(u)dF(u') + \int \Gamma^2(u)dF(u). \tag{12}$$

THEOREM 5.2. *Under assumptions A–D and* (5), *the statistic*

$$\widehat{R}_n(\Gamma) \stackrel{\text{def}}{=} \widehat{\Lambda}_{-2} - 2(1-nN^{-1})\int u^{-1}\Gamma(u)dF_n(u) + 2nN^{-1}\iint \frac{\Gamma(u)-\Gamma(u')}{u'-u}dF_n(u)dF_n(u') + \int \Gamma^2(u)dF_n(u) \tag{13}$$

is such that $\widehat{R}_n(\Gamma) \xrightarrow{2} R(\Gamma)$.

Proof. We start from Theorem 5.1. By Lemma 5.1,

$$\Lambda_{-1} = (1-y)\int u^{-1}dF(u),$$
$$\int (\Lambda_{-1} - (1-y)th(-t))d\eta(t) = (1-y)\int u^{-1}\Gamma(u)dF(u),$$
$$\int th^2(-t)d\eta(t) = \iint \frac{\Gamma(u)-\Gamma(u')}{u-u'}dF(u)dF(u'),$$

where the integrand is extended by continuity to $u = u'$. By Remark 6 we have

$$n^{-1}\text{tr } \Gamma^2(C) = \int \Gamma^2(u)dF_n(u) \xrightarrow{2} \int \Gamma^2(u)dF(u).$$

In view of (5) the moments $\int u^{-6}dF_n(u)$ exist and are uniformly bounded. This means that the contribution of small u to (13) can be made arbitrarily small. For $u > \varepsilon > 0$ the convergence $\widehat{R}_n(\Gamma) \xrightarrow{2} R(\Gamma)$ follows from Remarks 5 and 6. We conclude that the statement of Theorem 5.2 holds. □

Example 1. Let $\Gamma(C) = \alpha(I + tC)^{-1}$ be an estimator of Σ^{-1}, $\alpha > 0$. Then we have

$$R(\Gamma) = \Lambda_{-2} - 2\alpha(\Lambda_{-1} - th(-t)s(-t)) + \alpha^2\frac{d}{dt}(th(-t)).$$

The estimator of this function is

$$\widehat{R}_n(\Gamma) \stackrel{\text{def}}{=} \widehat{\Lambda}_{-2} - 2\alpha(n^{-1} - N^{-1}) \text{ tr } (C^{-1}(I+tC)^{-1}) \\ + 2\alpha n^{-1} N^{-1} \text{tr}^2 (I+tC)^{-1} + \alpha^2 n^{-1} \text{tr } (I+tC)^{-2}.$$

Now let us rewrite (12) in the form convenient for an explicit minimization. Denote

$$g_n(w) = w^{-1} h_n(w^{-1}) = \int (w-u)^{-1} dF_n(u).$$

If $c_1 > 0$ and $y > 0$, then $g_n(w) \xrightarrow{2} g(w)$. By Lemma 5.1, the function $g(w)$ satisfies the Hölder inequality, and for $u > 0$, we have Re $g(u) = \int_p (u-u')^{-1} dF(u')$, where we use the principle value of the integral.

Let us reduce (12) to the form allowing an explicit minimization. Denote

$$\Gamma^{\text{opt}}(u) = (1-y)u^{-1} + 2y \text{ Re } g(u).$$

Remark 7. Relation (12) can be rewritten as follows:

$$R(\Gamma) = R^{\text{opt}} + \int (\Gamma(u) - \Gamma^{\text{opt}}(u))^2 dF(u),$$
$$R^{\text{opt}} = \Lambda_{-2} - \int (\Gamma^{\text{opt}}(u))^2 dF(u).$$

Example 2. The equation (4) allows the solution in an analytical form for a two-parametric 'ρ-model' (see Chapter 2) of the limit distribution $F_0(u) = F_0(u, \sigma, \rho)$ of the eigenvalues of Σ with the limit density function

$$F_0'(u) = \begin{cases} (2\pi\rho)^{-1}(1-\rho)u^{-2}\sqrt{(c_2-u)(u-c_1)}, & u_1 \le u \le u_2, \\ 0 \quad \text{otherwise}, \end{cases}$$

where $u > 0$, $0 \le \rho < 1$, $c_1 = \sigma^2(1-\sqrt{\rho})^2$, $c_2 = \sigma^2(1+\sqrt{\rho})^2$, where $\sigma > 0$. For this model, the function $h(z)$ satisfies the equation $(1-h(z))(1-\rho h(z)) = \kappa z h(z) s(z)$, where $\kappa = \sigma^2(1-\rho)^2$ and $s(z) =$

$1 - y + yh(z)$. In this special case, the extremum solution can be found explicitly:

$$\Gamma^{\text{opt}}(u) = (1 - y + 2y \text{ Re } h(u^{-1}))u^{-1} = (\rho + y)(\rho u + \kappa y)^{-1}.$$

The equation

$$\Gamma^{\text{opt}}(u) = \int_{t \geq 0} (1 + ut)^{-1} d\eta^{\text{opt}}(t), \quad u \geq 0,$$

has the solution

$$\eta^{\text{opt}}(t) = \begin{cases} 0 & \text{for } t < \rho\kappa^{-1}y^{-1}, \\ (\rho + y)\kappa^{-1}y^{-1} & \text{for } t \geq \rho\kappa^{-1}y^{-1} \end{cases}$$

if $\rho > 0$. The estimator $\Gamma^{\text{opt}}(C) \in \mathfrak{K}^{(3)}$. Evaluating $R^{\text{opt}} = R^{\text{opt}}(\Gamma)$, we obtain $R^{\text{opt}} = \rho y(\rho + y)^{-1}\kappa^{-2}$. For special cases when $\rho = 0$ or $y = 0$, we obtain $R^{\text{opt}} = 0$.

Example 3. Consider the same model as in Example 2 but assume, in addition, that it is known *a priori* that the populations have the distribution functions $F_{0n}(u)$ of eigenvalues of Σ approaching $F_0(u) = F_0(u, \sigma, \rho)$. It then suffices to construct convergent estimators for the parameters σ^2 and ρ. Consider the statistics

$$\widehat{M}_\nu = n^{-1}\text{tr } C^\nu, \quad \nu = 1, 2, \quad \widehat{\sigma}^2 = \widehat{M}_2/\widehat{M}_1, \quad \widehat{\rho} = 1 - \widehat{M}_1^2/\widehat{M}_2.$$

For these, as $n \to \infty$, we have

$$\widehat{M}_1 \xrightarrow{2} M_1, \quad \widehat{M}_2 \xrightarrow{2} M_2, \quad \widehat{\sigma}^2 \xrightarrow{2} \sigma^2, \quad \text{and } \widehat{\rho} \xrightarrow{2} \rho,$$

where M_1 and M_2 are defined by (3). We suggest the following estimator of Σ^{-1}:

$$\widetilde{\Sigma}^{-1} = \widehat{\sigma}^2(\widehat{\rho} + nN^{-1})(\widehat{\rho}\,\widehat{\sigma}^2 C + n^{-1}N^{-1}\text{tr}^2 C \cdot I)^{-1}$$

The matrices $\widetilde{\Sigma}^{-1}$ only have eigenvalues that are uniformly bounded with the probability $p_n \to 1$ as $n \to \infty$. We can conclude that $\|\widetilde{\Sigma}^{-1} - \Gamma^{\text{opt}}(C)\| \to 0$ in probability. By Theorem 5.1,

$$\underset{n\to\infty}{\text{l.i.m.}}\; n^{-1}\text{tr } (\Sigma^{-1} - \widetilde{\Sigma}^{-1})^2 = R^{\text{opt}}.$$

Thus for populations with $F_{0n}(u) \to F_0(u, \sigma, \rho)$, the family of estimators $\{\widetilde{\Sigma}^{-1}\}$ have quadratic losses asymptotically not greater than the quadratic losses of any estimators from $\mathfrak{K}^{(3)}$.

Asymptotically Unimprovable Estimators of the Inverse Covariance Matrices

In the general case, the function $\widehat{R}_n(\Gamma)$ of the form (13) attains no minimum for any smooth $\Gamma(u)$. In the estimator

$$\Gamma_n^{\rm opt}(u) = (1 - nN^{-1})u^{-1} + 2nN^{-1}{\rm Re}\ g_n(u)$$

of the limit function $\Gamma^{\rm opt}(u)$, the function $g_n(u)$ is singular. To obtain a regular estimator, we introduce a smoothing of $g_n(u)$. We consider $g_n(w)$ and $\Gamma_n(w)$ for complex arguments w with Im $w = \varepsilon \neq 0$. First, we consider the regularized limit estimator

$$\Gamma_\varepsilon^{\rm opt}(u) = {\rm Re}\ [(1-y)w^{-1} + 2yg(w)],$$

where $w = u - i\varepsilon, \quad \varepsilon > 0$. For the estimator $\tilde{\Sigma}^{-1} = \Gamma(C)$, define the quadratic loss function

$$L_n = L_n(\Gamma) = n^{-1}{\rm tr}\ (\Sigma^{-1} - \Gamma(C))^2.$$

Let us prove the convergence of $L_n(\Gamma_\varepsilon^{\rm opt})$ to the same expression (12).

LEMMA 5.3. *Under assumptions A-D and* (5) *for* $\varepsilon > 0$,

$$L_n(\Gamma_\varepsilon^{\rm opt}) = R(\Gamma_\varepsilon^{\rm opt}) + \xi_n(\varepsilon) + r(\varepsilon),$$

where $\mathbf{E}\ |\xi_n(\varepsilon)|^2 \to 0$ *as* $n \to \infty$ *for fixed* $\varepsilon > 0$, *and* $r(\varepsilon) = O(\varepsilon)$ *as* $\varepsilon \to +0$.

Proof. For a fixed $\varepsilon > 0$, the function $\Gamma_\varepsilon^{\rm opt}(u)$ is continuous and has a bounded variation for $u > 0$. The value

$$L_n(\Gamma_\varepsilon^{\rm opt}) = n^{-1}{\rm tr}\ \Sigma^{-2} - 2n^{-1}{\rm tr}\ [\Sigma^{-1}\Gamma_\varepsilon^{\rm opt}(C)] + n^{-1}{\rm tr}\ [\Gamma_\varepsilon^{\rm opt}(C)]^2.$$

We note that $n^{-1}{\rm tr}\ \Sigma^{-2} \to \Lambda_{-2}$ in the right hand side, and

$$n^{-1}{\rm tr}\ [\Gamma_\varepsilon^{\rm opt}(C)]^2 = \int [\Gamma_\varepsilon^{\rm opt}(u)]^2 dF_n(u) \xrightarrow{2} \int [\Gamma_\varepsilon^{\rm opt}(u)]^2 dF(u).$$

Compare these expressions with (12). We note that it suffices to demonstrate the convergence

$$n^{-1}\text{tr }\Sigma^{-1}\varphi(C) \xrightarrow{2} \int (1-y)^{-1}u^{-1}\varphi(u)dF(u) \\ -y\iint \frac{\varphi(u)-\varphi(u')}{u'-u}dF(u)dF(u') \qquad (14)$$

for $\varphi(u) = \Gamma_\varepsilon^{\text{opt}}(u)$, where the last integrand is extended by continuity to $u = u'$. This relation is linear with respect to $\varphi(u)$, and it suffices to show (14) for $\varphi(u) = w^{-1}$ and $\varphi(u) = \text{Re } g_n(w)$, $w = u - i\varepsilon, \quad \varepsilon > 0$. As $n \to \infty$, we have

$$n^{-1}\text{tr }\left(\Sigma^{-1}(C - i\varepsilon I)^{-1}\right) \xrightarrow{2} (1-y)^{-1}\Lambda_{-2} + O(\varepsilon).$$

It follows that

$$(1-y)\int u^{-1}w^{-1}dF(u) - y\ [\int w^{-1}dF(u)]^2 \\ = (1-y)M_{-2} - yM_{-1}^2 + O(\varepsilon) = (1-y)^{-1}\Lambda_{-2} + O(\varepsilon).$$

For $\varphi(u) = g(w), \quad w = u - i\varepsilon, \quad \varepsilon > 0$, we find that

$$n^{-1}\text{tr }\left[\Sigma^{-1}g(C - i\varepsilon I)^{-1}\right] = \int n^{-1}\text{tr }\left[\Sigma^{-1}(C - z^{-1}I)^{-1}\right] dF(u) \\ = -\int zn^{-1}\text{tr }\left[\Sigma^{-1}(I - zC)^{-1}\right] dF(u),$$

where $z = (u - i\varepsilon)^{-1}$. By statements 2 and 5 of Lemma 5.1, for fixed $\varepsilon > 0$ and $u < u_2 = c_2(1+\sqrt{y})^2$ uniformly in u, we have

$$\mathbf{E}\ n^{-1}\text{tr }\left(\Sigma^{-1}(I - zC)^{-1}\right) = \\ = n^{-1}\text{tr }\Sigma^{-1} + zs(z)\mathbf{E}\ n^{-1}\text{tr }(I - zC)^{-1} + o_n'(\varepsilon) \\ = \Lambda_{-1} + zs(z)h(z) + o_n''(\varepsilon),$$

where $z = (u - i\varepsilon)^{-1}$, $o_n'(\varepsilon) \to 0$ and $o_n''(\varepsilon) \to 0$ as $n \to \infty$. In view of statement 3 of Lemma 5.1 for $\varepsilon \neq 0$ as $n \to \infty$, we obtain that var $[n^{-1}\text{tr }\left(\Sigma^{-1}(I - zC)^{-1}\right)] \to 0$ uniformly in u, and the expressions in (15) converges in the square mean. Substituting $g(z^{-1}) = zh(z)$ and $s(z) = 1 - y + yz^{-1}g(z^{-1})$, we find that the right hand side of (15) equals

$$-\Lambda_1 M_1 - (1-y)\int zg(z^{-1})dF(u) - y\int g^2(z^{-1})dF(u) + \xi_n(\varepsilon) + o_\varepsilon, \tag{16}$$

where $z = (u - i\varepsilon)^{-1}$, $o_\varepsilon \to 0$ as $\varepsilon \to +0$, and for fixed $\varepsilon > 0$, $\mathbf{E}\,|\xi_n(\varepsilon)|^2 \to 0$ as $n \to \infty$. We recall that the arguments u are bounded in the region of integration. In view of the Hölder condition for $g(z)$, the contribution to (16) of the difference between z and u^{-1} is of the order of magnitude of ε^ζ as $\varepsilon \to +0$ where $\zeta > 0$. We find that

$$2\int n^{-1}\text{Re}\; g(u)dF(u) = 2\int \left[\int_p (u-u')^{-1}dF(u')\right]u^{-1}dF(u) =$$
$$= -[\int u^{-1}dF(u)]^2 = -M_{-1}^2,$$

where $M_{-1} = (1-y)^{-1}\Lambda_{-1}$. It follows that (16) equals

$$(1-y)\int u^{-1}g^*(u)dF(u) - y\int g^2(u)dF(u) + o_\varepsilon + \xi_n(\varepsilon), \tag{17}$$

where $o_\varepsilon \to 0$ as $\varepsilon \to +0$. On the other hand, we substitute $\varphi(u) = g(w)$, where $w = u - i\varepsilon$, in the right hand side of (14). Using the identity

$$\iint \frac{g(u-i\varepsilon) - g(u'-i\varepsilon)}{u'-u}dF(u)dF(u') = \int g^2(u-i\varepsilon)dF(u), \tag{18}$$

we find that, for $\varphi(u) = g(w)$, $w = u - i\varepsilon$, the right hand side of (14) equals

$$(1-y)\int u^{-1}g(w)dF(u) - y\int g^2(w)dF(u). \tag{19}$$

In view of the Hölder condition, the real parts of (19) and (17) differ by $o_\varepsilon + \xi_n(\varepsilon)$. Thus the convergence (14) is proved for $\varphi(u) = g(w)$ and for $\varphi(u) = \text{Re}\; g(w)$, and $\varphi(u) = \Gamma_\varepsilon^{\text{opt}}(u)$ as well. The lemma's statement follows. □

For the regularized extremal function $\Gamma_{\varepsilon}^{\text{opt}}(\cdot)$, we can offer a 'natural estimator'

$$\Gamma_{n\varepsilon}^{\text{opt}}(u) = \text{Re } [(1 - nN^{-1})w^{-1} + 2nN^{-1}g_n(w)], \tag{20}$$

where $w = u - i\varepsilon$.

Remark 8. For $\varepsilon > 0$ and $u > 0$, the scalar function

$$\Gamma_{n\varepsilon}^{\text{opt}}(u) = \Gamma_{\varepsilon}^{\text{opt}}(u) + u^{-1}o_n + \xi_n(u, \varepsilon),$$

where $o_n \to 0$ and for any $u > 0$ $\mathbf{E}\ \xi_n^2(u,\varepsilon) \to 0$ as $n \to \infty$ for any $u > 0$. For $u \geq c_1 > 0$, the function $\Gamma_{\varepsilon}^{\text{opt}}(u) = \Gamma^{\text{opt}}(u) + r(\varepsilon)$, where $r(\varepsilon) = O(\varepsilon^{\zeta})$ as $\varepsilon \to +0$, $\zeta > 0$.

THEOREM 5.3. *Under conditions A–D and* (5),

$$\lim_{\varepsilon \to +0} \lim_{n \to \infty} \mathbf{E}\ |L_n(\Gamma_{n\varepsilon}^{\text{opt}}) - R(\Gamma^{\text{opt}})|^2 = 0.$$

Proof. The difference

$$L_n(\Gamma_{n\varepsilon}^{\text{opt}}) - L_n(\Gamma_{\varepsilon}^{\text{opt}}) = n^{-1}\text{tr } \left[Q(\Gamma_{n\varepsilon}^{\text{opt}}(C) - \Gamma_{\varepsilon}^{\text{opt}}(C))\right], \tag{21}$$

where $Q = -2\Sigma^{-1} + \Gamma_{n\varepsilon}^{\text{opt}}(C) + \Gamma_{\varepsilon}^{\text{opt}}(C)$. The eigenvalues of Q are bounded for $\varepsilon > 0$, and in view of Remark 8, the right hand side of (21) is not greater than $o(n^{-1})$ tr $C^{-1} + |\xi_n(u,\varepsilon)|$. Hence, $\mathbf{E}\ L_n(\Gamma_{n\varepsilon}^{\text{opt}})$ approaches $\mathbf{E}L_n(\Gamma_{\varepsilon}^{\text{opt}})$ in the square mean as $n \to \infty$. By Lemma 5.3 it follows that $\mathbf{E}\ |L_n(\Gamma_{\varepsilon}^{\text{opt}}) - R(\Gamma_{\varepsilon}^{\text{opt}})|^2 \xrightarrow{2} 0$. For $R(\cdot)$ from (12), consider the difference $d = R(\Gamma_{\varepsilon}^{\text{opt}}) - R(\Gamma^{\text{opt}})$ for small $\varepsilon > 0$. By Remark 8, the first two terms in (12) provide a contribution to d of the order of magnitude of ε^{ζ}. In view of (18) taking into account the Hölder condition for $g(z)$ we find that the third and the fourth term in (12) also contribute $O(\varepsilon^{\zeta})$ to d.2 The statement of our theorem follows. □

Corollary. Under assumptions A–D and (5), the family of estimators $\{\Gamma_{n\varepsilon}^{\text{opt}}(C)\}$ of the matrices Σ^{-1} defined by (20) is ε-dominating over the class $\mathfrak{K}^{(3)}$ with respect to the quadratic loss function $L_n(\cdot)$.