Can talk about why sample covariance matrix dispersed. In Ledoit 2001.

then 1977 Stein

preserves old eigenvectors. [are they good estimators?]

Loss function: still Stein's Loss.

~~Invariance: Orthogonal matrices.~~

Class of estimator we're considering: eigenvalue decomposition of $S = BLB^T$ $\quad \hat{\Sigma}(S) = B\varphi(L)B^T$ ~~st.~~

$$\underbrace{L_1 \geq L_2 \geq \cdots \geq L_p. \qquad \varphi_1(L) \geq \cdots \geq \varphi_p(L)}_{\frac{\varphi_1(L)}{\varphi_p(L)} << \frac{L_1}{L_p}.}$$

method: What's supposed to do: ~~$L(\Sigma, \hat{\Sigma}(S)) = tr\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det\Sigma^{-1}\hat{\Sigma}(S) - p.$~~

minimize Risk $\quad EL(\Sigma, \hat{\Sigma}(S)) = E\left[tr\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det\Sigma^{-1}\hat{\Sigma}(S) - p\right]$

what we're only capable of doing:

~~minimize Risk~~. - find an unbiased estimator of risk.

- minimize it.

the unbiased estimator:

Define $\Psi(L) = L^{-1}\varphi(L)$. $\Rightarrow \psi_i(L) = \frac{\varphi_i(L)}{L_i}$ $\qquad \psi_{jj} = \frac{\partial}{\partial L_j}\psi_j(L)$

$$E\left[(n-p+1)\sum\psi_j(L) - \sum\log\psi_j(L) + 2\sum_j\sum_{i>j}\frac{L_j\psi_j(L) - L_i\psi_i(L)}{L_j - L_i} + \underbrace{2\sum_j L_j\psi_{jj}(L)}_{\text{need to ignore to proceed.}} - \sum_j E\log\chi^2_{n-j+1} - p\right]$$

Final solution: $\varphi_j(L) = \frac{L_j}{\alpha_j(L)}$ $\qquad \alpha_j(L) = n+p-1+2L_j\sum_{i\neq j}\frac{1}{L_j - L_i}$

problem: α_j maybe negative.

$\frac{L_j}{\alpha_j}$ may not be ordered correctly.

solution: isotonizing algorithm.

~~L_1~~	~~α_1~~
~~L_2~~	
L_1	α_1
L_2	α_2
L_3	α_3
$\vdots$	$\vdots$
L_{p-4}	α_{p-4}
L_{p-3}	α_{p-3}
L_{p-2}	α_{p-2}
L_{p-1}	α_{p-1}
L_p	α_p

① merge negative α_j with α_{j-1} until all positive

② merge violent pairs until $\frac{L_j}{\alpha_j}$ are ordered correctly

③ averge ~~α~~ for each block.

~~ate~~ exactly the solution to an isotonic regression problem:

$$\sum_{j=1}^{p}\left[\left(\frac{L_j}{\alpha_j} - \varphi_j(L)\right)\right]^2\alpha_j \quad \text{subject to} \quad \varphi_1(L) \geq \cdots \geq \varphi_p(L).$$

No theoretical result since objective function too complicated.

Stein 1977.

$X_1, \cdots X_n$ iid $N(0_p, \Sigma_{p\times p})$ Σ non-singular. $S = XX^T \Rightarrow \hat{\Sigma} = \frac{1}{n}S$.

~~Loss~~

Loss function $L(\Sigma, \hat{\Sigma}) = tr(\Sigma^{-1}\hat{\Sigma}) - \log\det \Sigma^{-1}\hat{\Sigma} - p$ very easily derived from KL-divergence

$f \sim$ pdf $N(0, \Sigma)$ $KL(\hat{f}\|f)$

$\hat{f} \sim$ pdf $N(0, \hat{\Sigma})$.

properties: ① $L(\Sigma, \hat{\Sigma}) \geq 0$ $=0$ iff $\Sigma = \hat{\Sigma}$.

② convex about $\hat{\Sigma}$

③ invariate under any linear transformation (non-singular) $L(\Lambda\Sigma\Lambda^T, \Lambda\hat{\Sigma}\Lambda^T) = L(\Sigma, \hat{\Sigma})$

$\frac{1}{n}S = \frac{1}{n}BLB^T$ $\underline{\hat{\Sigma}(S) = B\varphi(L)B^T}$ proposed. $\frac{\varphi_1(L)}{\varphi_p(L)} << \frac{L_1}{L_p}$

Formulating the problem: minimize $L(\Sigma, \hat{\Sigma}(S)) = tr\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det \Sigma^{-1}\hat{\Sigma}(S) - p$.

to get an expression that we can deal with.

① need $\boxed{\text{lemma}}$: $Y \sim N(0,1)$. $E\,g'(Y) = E\,Y g(Y)$. $g: R \to R$ s.t. $E|g'(Y)| < +\infty$

$\cancel{\int g'(y) f(y) dy = \int f(y) d g(dy) = f(y)g(y)|_{-\infty}^{+\infty} - \int}$

$\int g'(y) f(y) dy = \int f(y)\, g(dy) = f(y)g(y)|_{-\infty}^{+\infty} - \int g(y) f'(y) dy$

$f(y) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2}) \Rightarrow f'(y) = -y\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2}) = -yf(y)$

$\Rightarrow \int g'(y) f(y) dy = \int y\, g(y) f(y) dy$

② differential operator for matrix. S_p: $\tilde{\nabla}_{ij} = \begin{cases} \frac{\partial}{\partial S_{ii}} \\ \frac{1}{2}\frac{\partial}{\partial S_{ij}} & i \neq j \end{cases}$

this can give us ~~dto~~ $\underbrace{dg(S)}_{\text{why scalar}} = \sum_j \sum_{i \leq j} \frac{\partial g(S)}{\partial S_{ij}} = tr[(\tilde{\nabla} g(S)) dS]$

③ $\cancel{E L(\Sigma, \hat{\Sigma}(S)) = E[tr\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det \Sigma^{-1}\hat{\Sigma}(S) - p]}$

Thm 2. $X \sim N(0, \Sigma)$ $S = XX^T$ $g: S_p \to S_p$ continuously differentiable

Then $E\,tr\,\Sigma^{-1}g(S) = E[n\,tr\,S^{-1}g(S) + 2tr\,S\,\hat{\nabla}(g(S)S^{-1})] = E[(n-p-1)tr\,S^{-1}g(S) + 2tr\,\tilde{\nabla} g(S)]$

Consider $h(S): S_p \to R_{p\times p}$. first assume $\Sigma = I$.

$E\,tr\,S h(S) = E\,tr\,\cancel{XX} XX^T h(S) = E\,tr\,X^T h(S) X = tr(E\,X^T h(S) X)$

i,j th element: $\sum_{k=1}^{p} E\,X_{ik} h_{ij}(S) X_{jk} = \sum_{k=1}^{p} E[X_{ik} X_{jk} h_{ij}(S) | X/X_{ik}]$

by lemma $= \sum_{k=1}^{p} \frac{d}{dX_{ik}} X_{jk} h_{ij}(S) = \delta_{ij} h_{ij}(S) + X_{jk}\, tr[(\tilde{\nabla} h(S)) \frac{dS}{dX_{ik}}]$

④ $EL(\Sigma,\hat{\Sigma}) = E\left[\mathrm{tr}\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det\Sigma^{-1}\hat{\Sigma}(S) - p\right]$

$$= E\Big[(n-p-1)\,\mathrm{tr}\,S^{-1}\hat{\Sigma}(S) + 2\,\mathrm{tr}\,\hat{\nabla}\,\hat{\Sigma}(S) \cancel{+\log\det - \log\det\hat{\Sigma}(S)}$$

$$+ \underbrace{\log\det S - \sum_{i=1}^{p} M\log\chi^2_{n-i+1}}_{\text{this is } \log\det\Sigma} - \log\det\hat{\Sigma}(S) - p\Big]$$

now $\hat{\Sigma}(S) = B\varphi(L)B^T$

$\Rightarrow EL(\Sigma,\hat{\Sigma}) = \cancel{E[(n-p-1)\,\mathrm{tr}\,B^{-1}L^{-1}B^TB\varphi(L)B^T}$

$$\overset{?}{=} \Big[(n-p-1)\,\mathrm{tr}\left[(BL^{-1}B^T)(B\varphi(L)B^T)\right] + 2\,\mathrm{tr}\,\tilde{\nabla}\,\hat{\Sigma}(S) + \log\det L - \sum_{i=1}^{p} E\log\chi^2_{n-i+1}$$

$$- \log\det\varphi(L) - p\Big] \quad \text{merge}$$

Define $\psi(L) = L^{-1}\varphi(L)$.

$$\Rightarrow\ = \Big[(n-p-1)\,\mathrm{tr}\,\psi(L) + 2\,\mathrm{tr}\,\hat{\nabla}\hat{\Sigma}(S) - \sum_{j=1}^{p}\log\psi_j(L) - \sum M\log\chi^2_{n-i+1} - p\Big]$$

expand $\cancel{\tilde{\nabla}\hat{\Sigma}(S) = \hat{\nabla}}$ $\hat{\nabla}(\hat{\Sigma})$

$\hat{\Sigma}(S) = B\varphi(L)B^T$? maybe consider $\hat{\nabla}_L\ \hat{\Sigma}(S)$.

remains a fucking mystery.

$B^T\hat{\Sigma}(S)B = \varphi(L) = L\psi(L)$.

$B^T\hat{\Sigma}(S)B\psi^{-1}(L) = L$.

$\hat{\Sigma}(S)B\psi^{-1}(L)B^T = S$.

$\Rightarrow \hat{\Sigma}(S) = B\psi(L)B^TS$

? $\hat{\nabla}\,\hat{\Sigma}(S) = \cancel{B\psi(L)B^T}$

Final version of error:

$$E\left[\mathrm{tr}\,\Sigma^{-1}\hat{\Sigma}(S) - \log\det\Sigma^{-1}\hat{\Sigma}(S) - p\right]$$

$$= E\Big[\underbrace{(n-p+1)\sum\psi_j(L) - \sum\log\psi_j(L) + 2\sum_j\sum_{i>j}\frac{L_j\psi_j(L) - L_i\psi_i(L)}{L_j - L_i} + 2\sum_j L_j\psi_{jj}(L) - \sum_j E\log\chi^2_{n-j+1} - p}_{\text{inner part is an unbiased (though I doubt) estimator of the error.}}\Big]$$

~~igno~~ ignoring term $2\sum_j L_j \psi_{jj}(L)$, take derivative over ψ_j

$$(n-p+1) - \frac{1}{\psi_j(L)} + 2\cdot\sum_{i>j}\frac{L_j}{L_j-L_i} - 2\sum_{i<j}\frac{L_j}{L_i-L_j} = 0$$

set $\alpha_j(L) = n+p-1+2L_j\sum_{i\neq j}\frac{1}{L_j-L_i}$

there is a "$-2j$" term in Stein's version. [but not in others] I have no idea

gives us $\psi_j(L) = \frac{1}{\alpha_j(L)} \Rightarrow \varphi_j(L) = \frac{L_j}{\alpha_j(L)}$.

Note: result is problematic. it can't retain order. sometimes can even be negative.

↓

Stein's Isotonizing Algorithm.