11-12.20

Factor Analysis (Amit Mitra : IIT Kanpur)

Let $\underline{X} \ni E\underline{X} = \underline{\mu}$, $\mathscr{C}(\underline{X}) = \Sigma$

Model: $\underline{X} - \underline{\mu} = L\,\underline{F} + \underline{\varepsilon}$

$\quad\quad\quad\quad\quad\quad\quad\quad\downarrow \quad \downarrow$

$\quad\quad\quad\quad\quad\quad\quad p\times m \quad m\times 1$

l_{ij} = looding of i^{th} vari. X_i on the j^{th} common factor - F_j

$\underline{F}$: $m\times 1$ unobservable RV called common factor.

l_{ij} indicates importance of X_i on F_j and is imp. for interpretation of F_j

$\underline{\varepsilon}$: called 'specific factor' specific to i^{th} vari.

Purpose : To explain cov. structure of RV. in terms of a few underlying but unobservable R.variable.

Assumptions : i) $\underline{F}$ is $\ni$ $E\underline{F} = \underline{0}$ & $\mathscr{C}(\underline{F}) = E(\underline{F}\underline{F}') = I_m$

ii) $Cov(\underline{\varepsilon}_{p\times 1}, \underline{F}_{m\times 1}) = E(\underline{\varepsilon}\underline{F}') = 0_{p\times m}$

iii) $\underline{\varepsilon}$ is $\ni$ $E\underline{\varepsilon} = 0$, $Cov(\underline{\varepsilon}) = diag(\psi_1, \psi_2, \cdots \psi_p)$
$= \Psi$

ψ_i, $i = 1, 2 \cdots p$ are called specific variances.

$\Sigma = Cov(\underline{X}) = E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'$

$= E(L\underline{F} + \underline{\varepsilon})(L\underline{F} + \underline{\varepsilon})'$

$= E(LFF'L') + LF\underline{\varepsilon}' + \underline{\varepsilon}F'L' + \underline{\varepsilon}\underline{\varepsilon}')$

$\Sigma = LL' + \Psi$

3

Now:

(1) $$X_i - \mu_i = \sum_{j=1}^{m} l_{ij} F_j + \varepsilon_i \quad , \quad i = 1, 2 \cdots p$$

$\mathrm{cov}(X_i, X_k)$

$$= \mathrm{cov}\left(\sum_{i=1}^{m} l_{ij} F_j + \varepsilon_i , \sum_{k=1}^{m} l_{kj} F_j + \varepsilon_k \right)$$

$$= l_{i1} l_{k1} + l_{i2} l_{k2} + \cdots + l_{im} l_{km}$$

$$= \sum_{j=1}^{m} l_{ij} l_{kj}$$

(2) $$\mathrm{cov}(\underline{X}, \underline{F}) = E\{(\underline{X} - \underline{\mu}) F'\}$$

$$= E\{(L\underline{F} + \underline{\varepsilon}) F'\}$$

$$= E(LFF') + E(\underline{\varepsilon} F')$$

$$= L\,E(FF') + \mathrm{cov}(\underline{\varepsilon}, \underline{F})$$

$$\mathrm{cov}(\underline{X}, \underline{F}) = L + \underline{0} \Rightarrow \mathrm{cov}(X_i, F_j) = l_{ij}$$

vect of com. factors..

(3) $$\mathrm{var}(X_i) = \mathrm{var}\left(\mu_i + \sum_{j=1}^{n} l_{ij} F_j + \varepsilon_i\right)$$

$$\sigma_{ii} = \sum_{j=1}^{m} l_{ij}^2 + \psi_i \qquad \because \mathrm{cov}(F_i, F_j) = 0, \; V(\varepsilon_i) = \psi_i$$

$$\sigma_{ii} = h_i^2 + \psi_i \qquad \text{where } h_i^2 = \sum_{j=1}^{m} l_{ij}^2$$

$= i^{th}$ communality

variability in X_i → m^{ith} communality $+ \psi_i$

contribution of m common factors $F_1, F_2, \cdots F_m$

Remark: 1

We say that an m-factor model holds for $\underline{X}$ if

$\underline{X}$ can be written as

$$\underline{X}-\underline{\mu} = L\underline{F}+\underline{\varepsilon}$$

Furthermore model for $\underline{X}$ holds iff Σ can be expressed as

$$\Sigma = LL' + \Psi$$

Example: Let $\underline{X}_{3\times1}$: vector of marks of students in maths, Phy & chem

& $\text{cov}(\underline{X}) = \begin{pmatrix} 1 & 0.83 & 0.78 \\ & 1 & 0.67 \\ & & 1 \end{pmatrix}$

Does a 1-factor model hold for $\underline{X}$?

$\longrightarrow$ $$\underline{X}-\underline{\mu} = L\underline{F}_{1\times1} + \underline{\varepsilon}$$

verify whether $\underset{}{\Sigma} = \underset{3\times1}{L}\,\underset{1\times3}{L'} + \underset{3\times3}{\Psi}$

$$L = \begin{pmatrix} l_{11} \\ l_{21} \\ l_{31} \end{pmatrix} \Rightarrow \Sigma = \begin{pmatrix} l_{11}^2+\Psi_1 & l_{11}l_{21} & l_{11}l_{31} \\ \underline{\text{Sym}} & l_{21}^2+\Psi_2 & l_{21}l_{31} \\ & & l_{31}^2+\Psi_3 \end{pmatrix}$$

$$\left.\begin{aligned} l_{11}l_{21} &= 0.83 \\ l_{11}l_{31} &= 0.78 \end{aligned}\right\} \Rightarrow \frac{l_{21}}{l_{31}} = \frac{0.83}{0.78}$$

$$l_{21}l_{31} = 0.67 \Rightarrow \frac{0.83}{0.78}l_{31}\,l_{31} = 0.67$$

$$\Rightarrow 0.67 \times \frac{0.78}{0.83} = l_{31}^2$$

$$\Rightarrow l_{31} = \pm 0.793$$

Take $l_{31} = 0.793$

Que 3

Q: 3

$$R^2 \text{ for } (X_2, X_3) \text{ model} = \frac{SS_{Res}(None) - SS_{Res}(X_2, X_3)}{SS_{Res}(None)}$$

$$= \frac{12.808 - 4.312}{12.808}$$

$$= 0.6633$$

$= 66.3\ \%$ → last option is correct

Que 4

		SS_{Res}
(X_1, X_3)	—	8.781
(X_1, X_4)	—	7.299
(X_2, X_3)	—	$\boxed{4.312}$ ✓ ∵ minimum
(X_3, X_4)	—	5.130

Choose model for which SS_{Res} is minimum ie R^2 will be maximum

Que 7

$$R^2_{adj} = 1 - \frac{(n-1)}{SS_{Res}(None)} \times \frac{SS_{Res}}{(n-p-1)}$$

$$= 1 - \frac{54-1}{12.808} \times \frac{7.4}{(54-1-1)}$$

$$= 1 - \frac{53}{12.808} \times \frac{7.4}{52}$$

$$= 0.411$$

n = no. of data pts = 54

model : $y = \beta_0 + \beta_4 x_4 + \varepsilon$

$\Rightarrow p = 1$

SS_{Res} for X_4 model = 7.4

Now $l_{31}^2 + \psi_3 = 1$

$\Rightarrow \quad \psi_3 = 1 - l_{31}^2 = 0.370$

= speci vari. of ith specific var.

Furth $\dfrac{l_{21}}{l_{31}} = \dfrac{0.83}{0.78}$

$\Rightarrow l_{21} = \dfrac{0.67}{l_{31}} = 0.845$ ②

$\Rightarrow \psi_2 = 1 - l_{21}^2 = 0.286$

$\Rightarrow l_{11} = \dfrac{0.83}{0.845} = 0.982$ ①

$\psi_1 = 1 - l_{11}^2 = 0.036$

$$L = \begin{pmatrix} 0.982 \\ 0.845 \\ 0.793 \end{pmatrix}, \quad \psi = \text{diag}\,(0.036, 0.286, 0.370)$$

$l_{11} = \text{corr}(X_1, F_1)$

$l_{12} = \text{corr}(X_2, F_1)$

$l_{13} = \text{corr}(X_3, F_1)$

$\therefore$ The 3 variables can be explained through 1 comm. factor. This comm factr. can be interpreted as the general ability.

$\therefore$ The resultant 1- fact model is

$$\underline{X} - \underline{\mu} = L F_{1\times 1} + \underline{\varepsilon}$$

$\text{corr}(X_1, F_1) = 0.982$;

Remark 2 : When $m = p$, Σ can always be written as $\Sigma = LL' + \Psi$ with $\Psi = 0$

This case is of no interest. $\because$ we are using p - comm. factors for p-variable.

Remark 3 Reduction in # of para. for an m-fact. model.

Σ has $\frac{p(p+1)}{2}$ distinct elements.

$\Sigma = LL' + \Psi$

$\left.\begin{array}{l} L \rightarrow pm \\ \Psi \rightarrow p \text{ entrie} \end{array}\right\}$ Total $pm + p$

Reduction in # of par.

$= \frac{p(p+1)}{2} - (pm + p)$

$= \frac{p}{2}(p+1 - 2m - 2)$

Ex: $p = 12$ and $m = 2$

Reduction $= \frac{12}{2}(12+1-4-2) = 42$

origin. par of $\Sigma = \frac{p(p+1)}{2} = \frac{\cancel{12}^{6} \times 13}{\cancel{2}} = 78$

Remark 4: Suppose m-fact. model holds for $\underline{x}$ and if $\underline{x}$ is rescaled ie

$$\underline{x} \Longrightarrow D\underline{x} = \underline{y}_{p\times 1}$$

where $D = \text{diag}(d_1, d_2, \cdots d_p)$.

Then m-factor model also holds for $\underline{y}$.

→ ∵ m-fact. model holds for $\underline{x}$,

$$\Rightarrow \underline{x} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$$

$$\Rightarrow D\underline{x} - D\underline{\mu} = DL\underline{F} + D\underline{\varepsilon}$$

$$\Rightarrow \underline{y} - \underline{\gamma} = (DL)\underline{F} + \underline{\eta} \qquad \underline{\eta} = D\underline{\varepsilon}, \quad \underline{\gamma} = D\underline{\mu}$$

$$\Rightarrow \underbrace{\underline{y} - \underline{\gamma}}_{p\times 1} = L^{*}\underline{F} + \underbrace{\underline{\eta}}_{p\times 1}$$

Here $E\underline{F} = \underline{0}$, $\mathscr{C}(\underline{F}) = I_m$; $E\underline{\eta} = \underline{0}$

$\text{cov}(\underline{\eta}) = D\Psi D'$

$\mathscr{C}(\underline{F}, \underline{\eta}) = \mathscr{C}(\underline{F}, D\underline{\varepsilon}) = E(\underline{F}\,\underline{\varepsilon}')D' = 0$

$\Rightarrow$ m-factor model holds for $\underline{y} = D\underline{x}$.

Remark 5: L and $\underline{F}$ in an m-factor model are not unique.

Suppose m-fact. model holds for $\underline{x}$.

$$\Rightarrow \underline{x} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$$

$$= L\Gamma\Gamma'\underline{F} + \underline{\varepsilon} \qquad \text{where } \Gamma\Gamma' = I$$

$$\Rightarrow \underline{x} - \underline{\mu} = (L)^{*}\underline{F}^{*} + \underline{\varepsilon} \qquad \text{where } L^{*} = L\Gamma$$

$$\underline{F}^{*} = \Gamma'\underline{F} \rightarrow (1)$$

Here $F^{*} = \Gamma'\underline{F}$ is ∋

$E(\underline{F}^{*}) = E(\Gamma'\underline{F}) = \Gamma' E(\underline{F}) = \underline{0}$

$\mathscr{C}(F^{*}) = \text{cov}(\Gamma'\underline{F}) = \Gamma' I_m \Gamma = I$

$\text{cov}(\underline{\varepsilon}, F^{*}) = \text{cov}(\underline{\varepsilon}, \Gamma'\underline{F}) = E(\underline{\varepsilon}F')\Gamma = 0$

$\Rightarrow$ (1) is another m-factor model for $\underline{x}$

Note: Some conditions are imposed so as to have the m-factor model unique.

eg. $L' \Psi^{-1} L$ to be a diagonal matrix.

Remark 6: Nonexistance of proper sol^n for m-factor model.

$\Sigma = LL' + \Psi$ be the m-factor that holds for $\underline{X}$.

i) if ψ_i's are negative then sol^n is not proper. Such a situation is said to 'Heywood' case.

Example of Heywood case

Let $\Sigma = \begin{pmatrix} 1 & 0.9 & 0.7 \\ & 1 & 0.4 \\ & & 1 \end{pmatrix}$

To check whether 1-factor model holds for $\underline{X}$

$\Sigma = LL' + \Psi$; $L = \begin{pmatrix} l_{11} \\ l_{21} \\ l_{31} \end{pmatrix}$, $\Psi = \begin{pmatrix} \psi_1 & 0 & 0 \\ & \psi_2 & 0 \\ & & \psi_3 \end{pmatrix}$

$$\Sigma = \begin{pmatrix} l_{11}^2 + \psi_1 & l_{11} l_{21} + 0 & l_{11} l_{31} \\ & l_{21}^2 + \psi_2 & l_{21} l_{31} \\ & & l_{31}^2 + \psi_3 \end{pmatrix}$$

$\Rightarrow l_{11}^2 + \psi_1 = 1$, $l_{11} l_{21} = 0.9$, $l_{11} l_{31} = 0.7$

$l_{21}^2 + \psi_2 = 1$, $l_{21} l_{31} = 0.4$

$l_{31}^2 + \psi_3 = 1$

$$\frac{l_{11}}{l_{21}} = \frac{0.7}{0.4} \Rightarrow l_{21} = \frac{0.4}{0.7} l_{11}$$

$$0.9 = l_{11} l_{21} \Rightarrow 0.9 = \frac{0.4}{0.7} l_{11}^2$$

$$\Rightarrow l_{11} = \pm 1.255 \rightarrow \text{Not feasible.}$$

Justification: Realize $V(X_1) = \sigma_{11} = 1 = V(F_1)$

$l_{11} = cov(X_1, F_1) = corr(X_1, F_1)$ $\quad \because V(X_1) = 1 = V(F_1)$

$\Rightarrow l_{11} = \pm 1.255$ is an absurd value.

Method of estimation of L and Ψ

Method I : Principal Comp. method.

Let (λ_i, e_i), $i=1,\dots,p$ be e-value, e-vector pairs

$\ni \lambda_1 \geq \lambda_2 \dots \geq \lambda_p \geq 0$ and

$\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ are the orthonormalized e-vectors.

$$\Rightarrow \Sigma = \lambda_1 \underline{e}_1 \underline{e}_1' + \dots + \lambda_p \underline{e}_p \underline{e}_p'$$

$$= \left(\sqrt{\lambda_1}\,\underline{e}_1 \quad \sqrt{\lambda_2}\,\underline{e}_2 \;\dots\; \sqrt{\lambda_p}\,\underline{e}_p\right) \begin{pmatrix} \sqrt{\lambda_1}\,\underline{e}_1' \\ \sqrt{\lambda_2}\,\underline{e}_2' \\ \vdots \\ \sqrt{\lambda_p}\,\underline{e}_p' \end{pmatrix}$$

$$= L L'$$

where $L = \left(\sqrt{\lambda_1}\,\underline{e}_1 \quad \sqrt{\lambda_2}\,\underline{e}_2 \dots \sqrt{\lambda_p}\,\underline{e}_p\right)$

Fits the factor model with as many factors as the number of variables with $\Psi = 0$: Null matrix.

We can get m-factor model with $m < p$.

Consider the situation where last $p-m$ e-values of Σ are negligible ie $\lambda_{m+1}, \dots, \lambda_p$ are close to zero

Thus $\Sigma \approx \lambda_1 \underline{e}_1 \underline{e}_1' + \dots + \lambda_m \underline{e}_m \underline{e}_m'$

$$\doteq \left(\sqrt{\lambda_1}\,\underline{e}_1 \dots \sqrt{\lambda_m}\,\underline{e}_m\right)_{p\times m} \begin{pmatrix} \sqrt{\lambda_1}\,\underline{e}_1' \\ \vdots \\ \sqrt{\lambda_m}\,\underline{e}_m' \end{pmatrix}_{m\times p} = LL'$$

$\therefore$ Variance of the specific factors can be taken as diagonal entries of $\Sigma - LL'$

ie $\psi_i = \sigma_{ii} - \sum_{j=1}^{m} l_{ij}^2 \qquad i=1,2,\dots,p$

$\therefore \Sigma \approx LL' + \Psi$ — approximation

Algorithm: Applying above procedure to given dataset :

Dataset : $[\underline{x}_1 \; \underline{x}_2 \dots \underline{x}_n]$

steps :

i) compute $\bar{\underline{x}}$, S : sample v-c-m.

ii) compute $(\hat{\lambda}_i, \hat{\underline{e}}_i)$, $i=1,2,\dots,p$ e-value-e-vector pairs of S

iii) The matrix of factor loading with m-vari. $(m<p)$ is estimated as

$$\hat{L} = [\sqrt{\hat{\lambda}_1}\,\hat{\underline{e}}_1 \;\dots\; \sqrt{\hat{\lambda}_m}\,\hat{\underline{e}}_m]$$

iv) The estimated specific variances $\hat{\psi}_i$ are given by diagonal entries of $S - \hat{L}\hat{L}'$

ie $\hat{\psi}_i = S_{ii} - \sum_{j=1}^{m} \hat{l}_{ij}^2$, $i = 1, 2 \cdots p$

$$\hat{\psi} = \text{diag}(\hat{\psi}_1, \hat{\psi}_2, \cdots \hat{\psi}_p)$$

v) The communalities are estimated as

$$\hat{h}_i^2 = \sum_{j=1}^{m} \hat{l}_{ij}^2$$

Note: i) For a principal comp. sol^n, the estimated factor loadings does not change as the number of factors are increased.

eg. For $m=1$, $\hat{L}_{(1)} = (\sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1)$

For $m=2$ $\hat{L}_{(2)} = [\sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1 \quad \sqrt{\hat{\lambda}_2}\, \hat{\underline{e}}_2]$

In general for $m=k$,

$$\hat{L}_{(k)} = [\sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1 \; \cdots \; \sqrt{\hat{\lambda}_k}\, \hat{\underline{e}}_k]$$

and for $m=k+1$, $\hat{L}_{(k+1)} = [\sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1, \cdots \sqrt{\hat{\lambda}_k}\, \hat{\underline{e}}_k \; \sqrt{\hat{\lambda}_{k+1}}\, \hat{\underline{e}}_{k+1}]$

Measure of approximation:

$$S \approx \hat{L}\hat{L}' + \psi$$

Result: If $\Delta = S - (\hat{L}\hat{L}' + \psi) = ((\delta_{ij}))$

then $\sum_{ij} \Delta_{ij}^2 = tr(\Delta^2) \leq \sum_{i=m+1}^{p} \hat{\lambda}_i^2$

Proof: The Δ matrix is $\ni$ $\quad \Delta_{ii} = 0$

ss of entries of $(S - \hat{L}\hat{L}' - \psi) \leq$ ss of entries of $(S - \hat{L}\hat{L}')$

$\Rightarrow \quad tr(\Delta^2) = \sum_i \sum_j \Delta_{ij}^2 \quad \leq \quad \pi$

$$S = (\sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1 \; \cdots \; \sqrt{\hat{\lambda}_p}\, \hat{e}_p) \begin{pmatrix} \sqrt{\hat{\lambda}_1}\, \hat{\underline{e}}_1 \\ \vdots \\ \sqrt{\hat{\lambda}_p}\, \hat{e}_p \end{pmatrix}$$

$$= \sum_{i=1}^{p} \hat{\lambda}_i \hat{\underline{e}}_i \hat{\underline{e}}_i'$$

$$= \hat{L}\hat{L}' + \sum_{j=m+1}^{p} \hat{\lambda}_j \hat{\underline{e}}_j \hat{\underline{e}}_j'$$

$$\Rightarrow S - \hat{L}\hat{L}' = \sum_{j=m+1}^{p} \hat{\lambda}_j \hat{\underline{e}}_j \hat{\underline{e}}_j'$$

$tr(\Delta^2) \leq tr(S - \hat{L}\hat{L}')^2$

$tr(S-\hat{L}\hat{L}')^2$

$= tr\left(\left[\sum_{j=m+1}^{p} \hat{\lambda}_j \hat{e}_j \hat{e}_j'\right]\left[\quad\right]\right)$

$= tr \sum_{j=m+1}^{p} (\hat{\lambda}_j \hat{e}_j \hat{e}_j')(\hat{\lambda}_j \hat{e}_j \hat{e}_j')$

$= tr \sum_{j=m+1}^{p} \hat{\lambda}_j^2 \hat{e}_j \hat{e}_j'$

$= \sum_{j=m+1}^{p} \hat{\lambda}_j^2 \, tr(\hat{e}_j \hat{e}_j')$

$= \sum_{j=m+1}^{p} \hat{\lambda}_j^2 \, tr(\hat{e}_j' e_j)$

$= \sum_{j=m+1}^{p} \hat{\lambda}_j^2 \cdot 1$

Thus $tr(\Delta^2) = \sum_{i,j} \Delta_{ij}^2 \leq \sum_{j=m+1}^{p} \hat{\lambda}_j^2$

Note: Contribution of factors to total sample variance.

Recall. $s_{ii} = \sum_{j=1}^{m} \hat{l}_{ij}^2 + \hat{\psi}_j$

- Contribution of first factor to $s_{ii} = \hat{l}_{i1}^2$

  Cont. — " — " — to total sample vari ($\because tr(S) = \sum_{i=1}^{p} s_{ii}$)

  $= (\hat{l}_{11}^2 + \hat{l}_{21}^2 + \cdots \hat{l}_{p1}^2)$

  $\hat{L} = (\sqrt{\hat{\lambda}_1}\, \hat{e}_1 \cdots \sqrt{\hat{\lambda}_m}\, \hat{e}_m)$

  where $\sqrt{\hat{\lambda}_j}\, \hat{e}_j = \begin{pmatrix} \hat{l}_{1j} \\ \hat{l}_{2j} \\ \hat{l}_{pj} \end{pmatrix}$

  $\sum_{i=1}^{p} \hat{l}_{ij}^2 = \hat{\lambda}_j \quad j = 1, 2 \ldots p$

- Proportion of TSV explained through 1st Fact $= \dfrac{\hat{\lambda}_1}{tr(S)}$

  $= \dfrac{\hat{\lambda}_1}{\sum_{i=1}^{m} s_i}$

method II Estimation of L and ψ

Maximum likelihood estimation

$$\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$$

$$\begin{pmatrix} \underline{F} \\ \underline{\varepsilon} \end{pmatrix} \sim N_{m+p}\left(\begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} I_m & 0 \\ 0 & \psi \end{pmatrix} \right)$$

$$\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon} = (L \quad I_p) \begin{pmatrix} \underline{F} \\ \underline{\varepsilon} \end{pmatrix}$$

$\underline{X} - \underline{\mu} \sim N_p(\underline{0}, \Sigma)$ $\quad \because \mathscr{E}\left((L \quad I_p)\begin{pmatrix} \underline{F} \\ \underline{\varepsilon} \end{pmatrix}\right)$

$$= (L \; I_p) \begin{pmatrix} I_m & 0 \\ 0 & \psi \end{pmatrix} \begin{pmatrix} L' \\ I_p \end{pmatrix}$$

$$= LL' + \psi = \Sigma$$

$\therefore$ LF is given by -

$$L(\underline{\mu}, \Sigma \mid \underline{x}_1, \cdots \underline{x}_n) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}A)$$
$$-\frac{n}{2}(\bar{\underline{x}} - \underline{\mu})'\Sigma^{-1}(\bar{\underline{x}} - \underline{\mu})$$

$$= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|LL' + \psi| - \frac{1}{2}tr\left((LL' + \psi)^{-1}A\right)$$
$$-\frac{n}{2}(\bar{\underline{x}} - \underline{\mu})'(LL' + \psi)^{-1}(\bar{\underline{x}} - \underline{\mu})$$

$\because$ choice of L is not unique we impose cond^n like

$L'\psi^{-1}L = A$ .. a diagonal matrix so that L is unique

$\therefore$ we maximize L s.t. $L'\psi^{-1}L = A$ gives mle of L & ψ with $\hat{\mu}_{MLE} = \bar{\underline{x}}$