c) Suppose the cases $(\underline{x}_i, Y_i)$ are iid from some pop and MLR 2 model $Y = \alpha + \underline{\beta}^T \underline{x} + e$ holds.

Then b) holds and

$$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 \Sigma_{\underline{x}}^{-1})$$

where $\underline{\beta} = \underline{\beta}_{OLS} = \Sigma_{\underline{x}}^{-1} \Sigma_{\underline{x}Y}$.

2] $\hat{\underline{\beta}} = \hat{\Sigma}_{\underline{x}}^{-1} \hat{\Sigma}_{\underline{x}Y}$. If e_i are iid, $\hat{\underline{\beta}} \xrightarrow{P} \underline{\beta}$, but it is possible that $\hat{\Sigma}_{\underline{x}}^{-1}$ and $\hat{\Sigma}_{\underline{x}Y}$ are not consistent estimators of $\Sigma_{\underline{x}}^{-1}$ and $\Sigma_{\underline{x}Y}$. It is possible that $\underline{\beta} \neq \Sigma_{\underline{x}}^{-1} \Sigma_{\underline{x}Y}$.

So (a) and (b) are much weaker than (c).

3] § 1.7 reviews the OLS MLR model at about the Math 484 level.

4) consider $\hat{\Sigma}_{\underline{x}}^{-1} \hat{\Sigma}_{\underline{x}Y}$ and $(X^T X)^{-1} X^T \underline{Y}$. If $p > n$, then generally $(X^T X)^{-1}$ and $\hat{\Sigma}_{\underline{x}}^{-1}$ do not exist.

common techniques: a) use $(X^T X + \lambda I_p)^{-1}$, $\lambda > 0$ RR

b) $(\text{diag } \hat{\Sigma}_{\underline{x}})^{-1} = \begin{pmatrix} \frac{1}{\hat{\sigma}_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\hat{\sigma}_p^2} \end{pmatrix}$, MMLE

c) Replace $\hat{\Sigma}_{\underline{x}}^{-1}$ by I_p, the $p \times p$ identity matrix. MMLE, OPLS

§3.9 OPLS paper online

5) The one component partial least squares (OPLS) estimator is easy to compute.

compute $\hat{\underset{\sim}{\Sigma}}_{\underline{x}Y} = \begin{pmatrix} \hat{\text{cov}}(x_1, Y) \\ \vdots \\ \hat{\text{cov}}(x_p, Y) \end{pmatrix} = \hat{\underline{\eta}}_{OPLS}$

Let $\overset{RV}{W} = \hat{A}\,\underline{x} = \hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}^T\, \underline{x}$. ← $\underbrace{\text{1 linear combo of the predictors}}$

Do the OLS regression of Y on W.

with working model (simple linear regression)

$$Y = \alpha + \lambda W + \varepsilon$$ to get

$\hat{\alpha}$ and $\hat{\lambda}$. Then

$$\hat{\alpha}_{OPLS} = \hat{\alpha} \quad \text{and} \quad \hat{\underline{\beta}}_{OPLS} = \hat{\lambda}\, \hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}$$

where $$\hat{\lambda} = \frac{\hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}^T \, \hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}}{\hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}^T \, \hat{\underset{\sim}{\Sigma}}_{\underline{x}} \, \hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}}.$$

Do not compute $\hat{\lambda}$ this way if p is large.

under iid cases, $\underline{\beta}_{OPLS} = \lambda \underset{\sim}{\Sigma}_{\underline{x}Y}$ with

$$\lambda = \frac{\underset{\sim}{\Sigma}_{\underline{x}Y}^T \, \underset{\sim}{\Sigma}_{\underline{x}Y}}{\underset{\sim}{\Sigma}_{\underline{x}Y}^T \, \underset{\sim}{\Sigma}_{\underline{x}} \, \underset{\sim}{\Sigma}_{\underline{x}Y}}.$$

6] The OPLS MLR model

$$Y = Y \mid \underline{\beta}_{OPLS}^T \underline{x} = \alpha_{OPLS} + \underbrace{\underline{\beta}_{OPLS}^T \underline{x}}_{\underline{x}^T \underline{\beta}_{OPLS}} + e.$$

7] The OLS regression of Y on $W = \hat{\underset{\sim}{\Sigma}}_{\underline{x}Y}^T \underline{x}$

e could use p

is $(\hat{\lambda}_{OLS} =)\ \hat{\lambda} = \hat{\Sigma}_{\underline{w}}^{-1} \hat{\Sigma}_{\underline{w}Y} =$

$$\hat{\Sigma}_{A\underline{x}}^{-1} \hat{\Sigma}_{A\underline{x},Y} = \left(\hat{\Sigma}_{\underline{x}Y}^{T} \hat{\Sigma}_{\underline{x}} \hat{\Sigma}_{\underline{x}Y}\right)^{-1} \hat{\Sigma}_{\underline{x}Y}^{T} \hat{\Sigma}_{\underline{x}Y}$$

$A = \hat{\Sigma}_{\underline{x}Y}^{T}$ (could use $\hat{A}$)

$$= \frac{\hat{\Sigma}_{\underline{x}Y}^{T} \hat{\Sigma}_{\underline{x}Y}}{\hat{\Sigma}_{\underline{x}Y}^{T} \hat{\Sigma}_{\underline{x}} \hat{\Sigma}_{\underline{x}Y}}.$$

8) Like the MCLT for $\overline{\underset{\sim}{Y}}$, the CLT for $\hat{\Sigma}_{\underline{x}Y}$ does not require matrix inversion.

9]* CLT for $\underset{\sim}{\Sigma}_{\underline{x}Y}$: Assume cases $(\underline{x}_i^T, Y_i)^T$ are iid. Assume $E(x_{ij}^k Y_i^m)$ exist for $j = 1, \ldots, p$ and $k, m = 0, 1, 2$.

Let $\underline{\mu}_x = E(\underline{x})$ and $\mu_Y = E(Y)$. Let "iid" $\underline{w}_i = (\underline{x}_i - \underline{\mu}_x)(Y_i - \mu_Y)$ with sample mean $\overline{\underline{w}}_n$. Let $\underline{\eta} = \underset{\sim}{\Sigma}_{\underline{x}Y} = \underset{\sim}{\eta}_{OPLS}$.

HD 12

a) $\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$,

$\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$,

$\sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$.

b) Let $\boldsymbol{z}_i = \boldsymbol{x}_i(Y_i - \overline{Y})$, $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})$.

Then $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$. Hence

$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$.

c) Let A be a full rank constant matrix with $k \le p$. Assume H_0 $A\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$ is true, and $\hat{\lambda} \xrightarrow{P} \lambda \neq 0$. Then

$\sqrt{n}\, A(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 A \boldsymbol{\Sigma}_{\boldsymbol{w}} A^T)$.

proof] a) $E(\boldsymbol{w}_i) = \boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ and $\text{Cov}(\boldsymbol{w}) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$.
The $\boldsymbol{w}_i$ are iid. Hence by the MCLT,

$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$.

HW2 answer

Now $n\hat{\underset{\sim}{\eta}}_n = \sum_i (\underline{x}_i - \bar{\underline{x}})(y_i - \bar{y}) =$

$$\sum_{i=1}^{n} (\underline{x}_i \overbrace{\underbrace{- \underline{\mu}_x}_{a} + \underbrace{\underline{\mu}_x - \bar{\underline{x}}}_{b}}^{0})(\underbrace{y_i \overbrace{- \mu_y}_{c} + \underbrace{\mu_y}^{0} - \bar{y}}_{d})$$

$$= \sum_i (\underbrace{\underline{x}_i - \underline{\mu}_x}_{a})(\underbrace{y_i - \mu_y}_{c}) +$$

$$\overbrace{\sum_i (\underbrace{\underline{x}_i - \underline{\mu}_x}_{a})(\underbrace{\mu_y - \bar{y}}_{d})}^{(n\bar{\underline{x}} - n\underline{\mu}_x)(\mu_y - \bar{y})} + (\underbrace{\underline{\mu}_x - \bar{\underline{x}}}_{b}) \overbrace{\sum_i (\underbrace{y_i - \mu_y}_{c})}^{\sum y_i - n\mu = n(\bar{y} - \mu)}$$

$$+ n(\underbrace{\underline{\mu}_x - \bar{\underline{x}}}_{b})(\underbrace{\mu_y - \bar{y}}_{d}) =$$

$$\sum_i \underset{\sim}{w}_i - n\underset{\sim}{a}_n - n\underset{\sim}{a}_n + n\underset{\sim}{a}_n =$$

$$\sum_i \underset{\sim}{w}_i - n\underbrace{(\underline{\mu}_x - \bar{\underline{x}})(\mu_y - \bar{y})}_{\underset{\sim}{a}_n} = n\hat{\underset{\sim}{\eta}}_n, \text{ Thus}$$

$$\hat{\underset{\sim}{\eta}}_n = \frac{1}{n}\sum_i \underset{\sim}{w}_i - \frac{\sqrt{n}(\bar{\underline{x}} - \underline{\mu}_x)\sqrt{n}(\bar{y} - \mu_y)}{n}$$

and $\sqrt{n}\,\hat{\underset{\sim}{\eta}}_n = \sqrt{n}\,\frac{1}{n}\sum_i \underset{\sim}{w}_i - \dfrac{\sqrt{n}(\bar{\underline{x}} - \underline{\mu}_x)\sqrt{n}(\bar{y} - \mu_y)}{\sqrt{n}}$.

$$\therefore \sqrt{n}(\hat{\underset{\sim}{\eta}}_n - \underset{\sim}{\eta}) = \sqrt{n}(\bar{\underline{w}}_n - \underline{\eta}) + \underbrace{\frac{\sqrt{n}(\bar{\underline{x}} - \underline{\mu}_x)\sqrt{n}(\bar{y} - \mu_y)}{\sqrt{n}}}_{\xrightarrow{P} 0}$$

$$= \sqrt{n}(\bar{\underline{w}}_n - \underset{\sim}{\eta}) + O_p(1)$$

$$\xrightarrow{D} N_p(\underset{\sim}{0}, \Sigma_{\underline{w}}) + \underset{\sim}{0} \sim N_p(\underset{\sim}{0}, \Sigma_{\underline{w}}).$$

Now $\sqrt{n}(\underline{\hat{\eta}} - \underline{\eta}) = \sqrt{n}\left(\frac{n}{n-1}\tilde{\underset{\sim}{\eta}} - \underset{\sim}{\eta}\right) =$

$$\sqrt{n}\left(\frac{n}{n-1}\tilde{\underset{\sim}{\eta}} \underbrace{- \frac{n}{n-1}\underline{\eta} + \frac{n}{n-1}\underset{\sim}{\eta}}_{0} - \underset{\sim}{\eta}\right)$$

$$\underset{\sim}{\eta} = \frac{(n-1)\underline{\eta}}{n-1}$$

$$= \sqrt{n}\,\frac{n}{n-1}(\tilde{\underset{\sim}{\eta}} - \underline{\eta}) + \sqrt{n}\left(\frac{\underset{\sim}{\eta}}{n-1}\right) =$$

$$\underbrace{\frac{n}{n-1}}_{\xrightarrow{P} 1} \sqrt{n}(\tilde{\underline{\eta}} - \eta) + \underbrace{\sqrt{n}\left(\frac{\eta}{n-1}\right)}_{\xrightarrow{P} \underset{\sim}{0}} \xrightarrow{D} N_p(\underset{\sim}{0}, \Sigma_{\underline{w}}).$$

b) $\sum_i \underline{w}_i = \sum_i (\underbrace{\underline{x}_i - \overline{\underline{x}}}_{a} + \underbrace{\overline{\underline{x}} - \underline{\mu}_x}_{b})(\underbrace{y_i - \overline{y}}_{c} + \underbrace{\overline{y} - \mu_y}_{d})$

$$= \sum_i (\underbrace{\underline{x}_i - \overline{\underline{x}}}_{a})(\underbrace{y_i - \overline{y}}_{b}) +$$

$$\overbrace{\sum_i (\underbrace{\underline{x}_i - \overline{\underline{x}}}_{a})(\underbrace{\overline{y} - \mu_y}_{d})}^{0} + (\underbrace{\overline{\underline{x}} - \underline{\mu}_x}_{b}) \underbrace{\sum_i (y_i - \overline{y})}_{0\ c} + n(\underbrace{\overline{\underline{x}} - \underline{\mu}_x}_{b})(\underbrace{\overline{y} - \mu_y}_{d})$$

$$= \sum_i \underline{z}_i + n(\overline{\underline{x}} - \underline{\mu}_x)(\overline{y} - \mu_y) = \sum_i \underset{\sim}{z}_i + n\,\underline{a}_n =$$

$\sum_i \underline{z}_i = \sum_i \underline{w}_i$ ← constant wrp i

$\sum_i (\underline{z}_i + \underline{a}_n) = \sum_i \underline{w}_i$. Hence

$$\sum_i (\underline{w}_i - \overline{\underline{w}})(\underline{w}_i - \overline{\underline{w}})^T = \sum_i (\underline{z}_i + \underline{a}_n - (\overline{\underline{z}}_n + \underline{a}_n))(\underline{z}_i - \overline{\underline{z}}_n)^T$$

$= \sum_i (\underline{z}_i - \bar{\underline{z}})(\underline{z}_i - \bar{\underline{z}})^T$.

Thus $\hat{\underline{\Sigma}}_{\underline{w}} = \hat{\underline{\Sigma}}_{\underline{z}} = \frac{1}{n-1} \sum_i (\underline{z}_i - \bar{\underline{z}})(\underline{z}_i - \bar{\underline{z}})^T$

and $\tilde{\underline{\Sigma}}_{\underline{w}} = \tilde{\underline{\Sigma}}_{\underline{z}} = \frac{n-1}{n} \hat{\underline{\Sigma}}_{\underline{z}}$.

c) If Ho is true then $A \underset{\sim}{\eta} = \underset{\sim}{0}$.

Hence $\sqrt{n}\, A(\hat{\underset{\sim}{\eta}} - \underset{\sim}{\eta}) = \sqrt{n}\, A \hat{\underset{\sim}{\eta}} \xrightarrow{D} N_k(\underset{\sim}{0}, A \underline{\Sigma}_{\underline{w}} A^T)$

and $\sqrt{n}\, \hat{\lambda} A \hat{\underset{\sim}{\eta}} = \sqrt{n}\, A(\hat{\lambda} \hat{\underset{\sim}{\eta}} - \underbrace{\lambda \underset{\sim}{\eta}}_{\lambda A \eta \overset{Ho}{=} 0}) \xrightarrow{D} N_k(\underset{\sim}{0}, \lambda^2 A \underline{\Sigma}_{\underline{w}} A^T)$.

p1.5

10) The proof used <u>Slutsky's Th</u>!

Let $\underset{\sim}{x}_n$ and $\underset{\sim}{y}_n$ be sequences of random vectors.
Let W_n be a sequence of $k \times k$ nonsingular random matrices. Let G be a $k \times k$ constant nonsingular matrix. Let $\underset{\sim}{x}_n \xrightarrow{D} \underset{\sim}{x}$ and $\underset{\sim}{y}_n \xrightarrow{P} \underset{\sim}{d}$ for some constant vector $\underset{\sim}{d}$.

i) $\underset{\sim}{x}_n + \underset{\sim}{y}_n \xrightarrow{D} \underset{\sim}{x} + \underset{\sim}{d}$

ii) $\underset{\sim}{y}_n^T \underset{\sim}{x}_n \xrightarrow{D} \underset{\sim}{d}^T \underset{\sim}{x}$

iii) If $W_n \xrightarrow{P} G$, then $W_n \underset{\sim}{x}_n \xrightarrow{D} G \underset{\sim}{x}$,

$$\underset{\sim}{x}_n^T W_n \xrightarrow{D} \underset{\sim}{x}^T C$$

$$W_n^{-1} \underset{\sim}{x}_n \xrightarrow{D} C^{-1} \underset{\sim}{x},$$ and

$$\underset{\sim}{x}_n^T W_n^{-1} \xrightarrow{D} \underset{\sim}{x}^T C^{-1}.$$

11) If $\underset{\sim}{x}_n \xrightarrow{P} \underset{\sim}{0}$ we say $x_n = o_p(1)$.

If $\underset{\sim}{x}_n \xrightarrow{D} N_k(\underset{\sim}{0}, \Sigma)$, we say $\underset{\sim}{x}_n = O_p(1)$.

$O_p(1) + o_p(1) = O_p(1)$. $O_p(1)\, o_p(1) = o_p(1)$.

12) when p is fixed, we often test

$H_0\ \underset{\sim}{\beta} = \underset{\sim}{0}$ with $A = I_p$

$$H_0 \begin{pmatrix} \beta_{i_1} \\ \vdots \\ \beta_{i_k} \end{pmatrix} = \underset{\sim}{0} \quad \text{with } A = \begin{bmatrix} 0 \cdots 0\ 1\ 0 \cdots 0 \\ 0 \cdots 0\ 1\ 0 \cdots\ 0 \\ \vdots \\ 0 \cdots 0 \cdots 0\ 1\ 0 \cdots 0 \end{bmatrix}$$

$\leftarrow i_1$ position

$\uparrow i_k$ position

1 in the i_j position

$$H_0\ \beta_i = 0 \quad \text{with } A = [0 \cdots 0\ 1\ 0 \cdots 0]$$

$\uparrow$ ith position.

13) In high dimensions, if $n \ge 10k$,

we can test $\underset{q \times k}{B} \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{ik} \end{pmatrix} = \underset{\sim}{0} = A\underset{\sim}{\beta}$.

$q \le k$

If $\lambda \ne 0$, then $A \beta_{OPLS} = \underset{\sim}{0}$ iff

$A \lambda \Sigma_{\underline{x}Y} = \underset{\sim}{0}$ iff $A \Sigma_{\underline{x}Y} = \underset{\sim}{0} = A\underset{\sim}{\eta}$.

So test $B \Sigma_{\underline{w},Y} = \underset{\sim}{0}$ with $\underline{w} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix}$.

In particular $H_0\ \beta_i = 0$ iff $H_0\ \eta_i = \text{cov}(x_i, Y) = 0$.

we can get a $100(1-\alpha)\%$ confidence (95% common)

interval (CI) for $\eta_i = \text{cov}(x_i, Y)$.

$\hat{\underset{\sim}{\eta}}_i \pm \underbrace{1.96}_{} \dfrac{\hat{\underset{\sim}{\Sigma}}_{\underline{w}ii}}{\sqrt{n}}$ for a 95% CI.

could replace $z_{.95} = 1.96$ by $t_{.95, n-a}$

(eg $a = 0, 1$ etc).

14) In high dimensions, often $\hat{\lambda}$ will

not be a good estimator of λ.

15] For iid cases where cov's exist and $\Sigma_{xy} \neq 0$,

$$\beta_{OPLS} = \lambda \Sigma_{xy} = \Sigma_x^{-1} \Sigma_{xy} = \beta_{OLS}$$

iff Σ_{xy} is an eigenvector of Σ_x^{-1} with eigenvalue λ iff Σ_{xy} is an eigenvector of Σ_x with eigenvalue $\frac{1}{\lambda}$:

$$\lambda \Sigma_x \Sigma_{xy} = \Sigma_{xy} \quad \text{so}$$

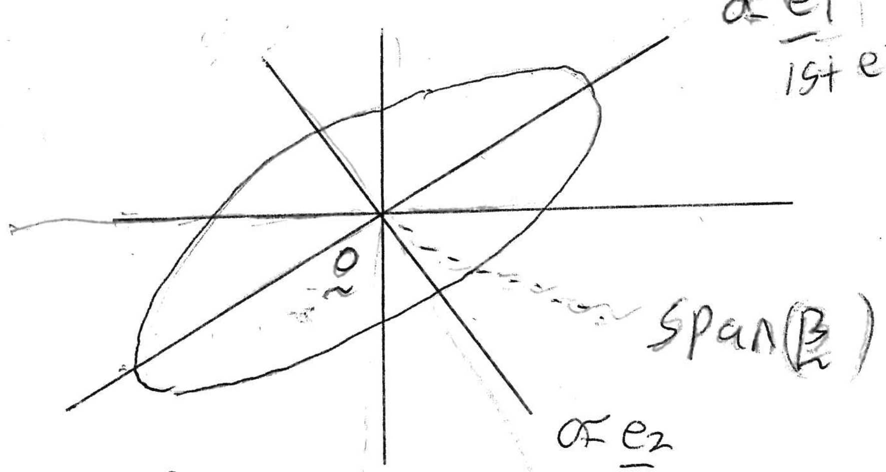
$$\Sigma_x \Sigma_{xy} = \frac{1}{\lambda} \Sigma_{xy}.$$

If $\Sigma_{xy} = 0$, then $\beta_{OPLS} = \beta_{OLS} = 0$.

If there is only one predictor ($p=1$), then $\beta_{OPLS} = \beta_{OLS}$.

If $\Sigma_x = I_p$, $\beta_{OPLS} = \beta_{OLS}$.

16) In general, $\hat{\underline{\beta}}_{OPLS}$ estimates $\underline{\beta}_{OPLS} \neq \underset{\sim}{\beta}_{OLS}$. The regularity condition $\underline{\beta}_{OLS} = \underset{\sim}{\beta}_{OPLS}$ is very strong.



Even with 2 predictors, $\underline{\beta}_{OLS} = \underline{\beta}_{OPLS} = c\,\underline{e}_1$ or $d\,\underset{\sim}{e}_2$. OLS theory says $\underset{\sim}{\beta}$ can be in the span of any vector in $\mathbb{R}^p$, not just in the span of an axis determined by the eigenvectors of cov($\underline{x}$). Linear combinations of the axes "$\underset{\sim}{e}_i$" can approx $\underset{\sim}{\beta}$, but the best single axis gives bad approx's.