Winter 14 – AMS225 Homework 3 Solution

1. (a) By differentiation, we see that for a fixed $B$, an optimal $\mu$ must satisfy

$$0 = A^T A (\mu_Y - \mu - B^T \mu_X).$$

This equation is satisfied if $\mu = \mu_Y - B^T \mu_X$. Substituting into the objective function we have that an optimal $B$ minimizes

$$E\|Z[(Y - \mu_Y) - B^T (X - \mu_X)]\|_2^2.$$

Differentiating once more, we have that an optimal $B$ satisfies

$$0 = (\Sigma_{XY} - \Sigma_{XX} B) A^T A.$$

This equation is satisfied if $B = \Sigma_{XX}^{-1} \Sigma_{XY}$. Looking back, we see that $(\mu_*, B_*)$ satisfy the two equations above, so they minimize

$$E\|A(Y - \mu - B^T X)\|_2^2.$$

However, they may not be unique minimizers.

(b) From the previous part, the optimal value of $E\|A(Y - \mu - B^T X)\|_2^2$ is obtained by $(\mu_*, B_*)$ and

$$E\|A(Y - \mu_* - B_*^T X)\|_2^2 = \text{Tr}\{A(\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}) A^T\}$$

Since it is optimal for all $A \geq 0$,

$$\text{Tr}\{A(\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}) A^T\} \leq E\|A(Y - \mu - B^T X)\|_2^2 \leq E\text{Tr}\{A(Y - \mu_* - B_*^T X)(Y - \mu_* - B_*^T X)^T A^T\}$$

for all $A \geq 0$. Now let $A = aa^T$, where $a \in \mathbb{R}^q$ and $\|a\|^2 = 1$. Then $aa^T aa^T = aa^T$, $a^T a = 1$ and the above inequality implies

$$a^T (\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}) a \leq a^T [E(\mu_Y - \mu - B^T \mu_X)(\mu_Y - \mu - B^T \mu_X)^T] a.$$

This is true for all unit vectors $a$, and, by homogeneity, it remains true if we allow $\|a\|^2 = 1$. So the above inequality implies

$$\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \leq E(\mu_Y - \mu - B^T \mu_X)(\mu_Y - \mu - B^T \mu_X)^T.$$

(c) First, we will show that if $C \geq D$ and $D > 0$, then $\det(C) \geq \det(D)$.

Proof. If $C \geq D$, then $x^T (C - D) x \geq 0$ for all $x \in \mathbb{R}^q$. Since $D$ is non-singular, we may take $x = D^{-1/2} y$ for some $y \in \mathbb{R}^q$ and conclude that

$$y^T (D^{-1/2} C D^{-1/2} - I_q) y \geq 0$$

for all $y \in \mathbb{R}^q$. Of course, this means that the eigenvalues of $D^{-1/2} C D^{-1/2}$ are no less than 1 and so

$$1 \leq \det(D^{-1/2} C D^{-1/2}) = \det(D^{-1}) \det(C).$$
$D > 0$ by assumption, so $\det(D^{-1}) > 0$ and the above inequality implies $\det(D) \leq \det(C)$.

Now specializing to our particular problem, we take $C = E(Y - \mu - B^T X)(Y - \mu - B^T X)^T$, $D = (\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY})$ and verify that $D > 0$, because it is the Schur complement within a positive definite matrix. The result of the previous part is that $C \geq D$.

2. (a)

\[
\begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix} \sim \text{MVN}_{2p}\left(\begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix}, (X^T X)^{-1} \otimes \Sigma_\epsilon\right).
\]

(b) Define

\[
\hat{\delta} = (I_p \otimes [1, -1]) \begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix}.
\]

Then our $\hat{\delta}$ follows multivariate Normal distribution with

\[
E(\hat{\delta}) = \delta = b_1 - b_2,
\]

and

\[
\text{Var}(\hat{\delta}) = (I_p \otimes [1, -1])\{(X^T X)^{-1} \otimes \Sigma_\epsilon\}(I_p \otimes [1, -1])^T
\]

\[
= (I_p \otimes [1, -1])\{(X^T X)^{-1} \otimes \Sigma_\epsilon\}(I_p \otimes \begin{bmatrix}1 \\ -1\end{bmatrix})
\]

\[
= (X^T X)^{-1}(\sigma_{11} + \sigma_{22} - 2\sigma_{12}),
\]

where

\[
\Sigma_\epsilon = \begin{bmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{12} & \sigma_{22}
\end{bmatrix}.
\]

(c) Let

\[
\frac{(\hat{\delta} - \delta)^T (X^T X)(\hat{\delta} - \delta)}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}} \sim \chi_q^2.
\]

We know that $nS_n \sim \mathcal{W}_2(\Sigma_\epsilon, n - q)$ that is independent of $\hat{\delta}$.

\[
n \frac{\hat{\sigma}_{11} + \hat{\sigma}_{22} - 2\hat{\sigma}_{12}}{\hat{\sigma}_{11} + \hat{\sigma}_{22} - 2\hat{\sigma}_{12}} \sim \mathcal{W}_1(1, n - q) \overset{d}{\sim} \chi_{n-q}^2,
\]

where

\[
S_n = \begin{bmatrix}
\hat{\sigma}_{11} \\
\hat{\sigma}_{12} \\
\hat{\sigma}_{12} \\
\hat{\sigma}_{22}
\end{bmatrix}.
\]

So, we have

\[
\frac{n - q}{q} \frac{(\hat{\delta} - \delta)^T (X^T X)(\hat{\delta} - \delta)}{\hat{\sigma}_{11} + \hat{\sigma}_{22} - 2\hat{\sigma}_{12}} \sim F_{q, n-q}.
\]
Now we invert this quantity to get a confidence region for $\delta$. The set
\[
\left\{ \delta : \frac{(\tilde{\delta} - \delta)^T (X^T X)(\tilde{\delta} - \delta)}{\hat{\sigma}_{11} + \hat{\sigma}_{22} - 2\hat{\sigma}_{12}} \leq \frac{q}{n-q} F_{q,n-q}(\alpha) \right\}
\]
is a $(1 - \alpha) \times 100\%$ confidence region for $b_1 - b_2$.

3. The column span of the design matrix includes the vector of 1s, because each row of $X$ sums to 1. So we should not include an (extra) intercept in this regression. Using R, we can compute the least squares fits by the commands:

\[
\begin{align*}
\text{> library(MMST)} \\
\text{> data(root.stocks)} \\
\text{> obj <- lm(cbind(Y1, Y2, Y3, Y4) ~ 0 + type, data = root.stocks)}
\end{align*}
\]

(a) The estimated regression coefficient matrix is
\[
\begin{array}{c}
\text{> obj} \\
\text{Call:} \\
\text{lm(formula = cbind(Y1, Y2, Y3, Y4) ~ 0 + type, data = root.stocks)}
\end{array}
\]

\[
\begin{array}{c}
\text{Coefficients:} \\
\text{Y1} & \text{Y2} & \text{Y3} & \text{Y4} \\
\text{typeI} & 113.75 & 2977.12 & 373.88 & 871.12 \\
\text{typeII} & 115.75 & 3109.12 & 451.50 & 1280.50 \\
\text{typeIII} & 110.75 & 2815.25 & 445.50 & 1391.37 \\
\text{typeIV} & 109.75 & 2879.75 & 390.62 & 1039.00 \\
\text{typeIX} & 85.87 & 1675.00 & 235.62 & 369.87 \\
\text{typeV} & 108.00 & 2557.25 & 431.25 & 1181.00 \\
\text{typeVI} & 103.62 & 2214.62 & 359.62 & 735.00 \\
\text{typeVII} & 105.25 & 2899.37 & 325.00 & 659.00 \\
\text{typeX} & 114.75 & 2998.25 & 371.12 & 841.12 \\
\text{typeXII} & 135.62 & 4699.37 & 477.25 & 1897.37 \\
\text{typeXIII} & 113.87 & 3055.50 & 391.50 & 1081.37 \\
\text{typeXV} & 122.00 & 3628.88 & 448.37 & 1289.25 \\
\text{typeXVI} & 121.37 & 4047.25 & 450.75 & 1695.62
\end{array}
\]

(b) An unbiased estimate of the error covariance matrix is \([104/(10413)]S_n:\)
\[
\begin{array}{c}
\text{> round(crossprod(resid(obj)) / (104-13), digits=1)} \\
\text{Y1} & \text{Y2} & \text{Y3} & \text{Y4} \\
\text{Y1} & 64.6 & 4100.9 & 164.7 & 814.6 \\
\text{Y2} & 4100.9 & 356106.1 & 13794.6 & 75982.3 \\
\text{Y3} & 164.7 & 13794.6 & 1127.5 & 6565.0 \\
\text{Y4} & 814.6 & 75982.3 & 6565.0 & 47838.2
\end{array}
\]

4. We estimate the standard errors by the square root of the diagonal of \((X^T X)^{-1} \otimes \Sigma_e\). Since the design is orthogonal, i.e. \(X^T X = 8I_4\), the estimated standard errors are the equal for each coefficient within each variate:
> sqrt(diag((crossprod(resid(obj)) / (104-13))/8))

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>Y2</td>
<td>Y3</td>
<td>Y4</td>
</tr>
<tr>
<td>2.842083</td>
<td>210.981666</td>
<td>11.871782</td>
<td>77.329000</td>
</tr>
</tbody>
</table>

Alternatively, you could use `summary(obj) or sqrt(diag(vcov(obj)))` to get the estimated standard errors.