Solutions for Econometrics I Homework No.1

due 2006-02-20

Feldkircher, Forstner, Ghoddusi, Grafenhofer, Pichler, Reiss, Yan, Zeugner

Exercise 1.1

Structural form of the problem:

1. $q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + u_{t1}$

2. $q_t^s = \beta_0 + \beta_1 p_{t-1} + u_{t2}$

   To get the reduced form solve your system of equations for the endogenous variables:

3. $q_t^s = q_t^d = \beta_0 + \beta_1 p_{t-1} + u_t$

4. $p_t = \frac{1}{\alpha_1}[(\beta_0 - \alpha_0) - \alpha_2 y_t + \beta_1 p_{t-1}(u_{t2} - u_{t1})]$

To arrive at the final form, each equation may only contain own lags or exogenous variables on the right-hand side. So (4) is already in the final form and for (3):

Rewrite (1) to get

$$p_t = \frac{1}{\alpha_1}[q_t^d - \alpha_0 - \alpha_2 y_t - u_{t1}]$$

If we lag this we get

$$p_{t-1} = \frac{1}{\alpha_1}[q_{t-1}^d - \alpha_0 - \alpha_2 y_{t-1} - u_{t1-1}]$$

Plug this into (3) to get

$$q_t^s = q_t^d = \beta_0 + u_t + \beta_1[\frac{1}{\alpha_1}(q_{t-1}^d - \alpha_0 - \alpha_2 y_{t-1} - u_{t1-1})]$$

Exercise 1.2

The variance covariance matrix (VCV) of $X \in \mathbf{R}^k$ is defined as $Var(X) = \mathbf{E}[(X - \mathbf{E}X)(X - \mathbf{E}X)']$.The covariance matrix is given by $Cov(X,Y) = \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)']$.Show the following transformation rules, where A, B, a, b are non-random matrices or vectors of suitabel dimensions ($A \in \mathbf{R}^{s\times k}, B \in \mathbf{R}^{t\times m}$ and $a \in \mathbf{R}^s, b \in \mathbf{R}^t$)

1. $\mathbf{E}(AX + a) = A\mathbf{E}(X) + a$
2. $Cov(X,Y) = \mathbf{E}(XY') - (\mathbf{E}X)(\mathbf{E}Y)'$
3. $Cov(AX + a, BY + b) = A[Cov(X,Y)]B'$
4. $Var(AX + a) = A[Var(X)]A'$

Proof:

1.

$$\mathbf{E}(AX + a) = A\mathbf{E}(X) + a \tag{1}$$

Martin Wagners Comment on that: "follows from properties of the integral" Dominikis comment: "multiply out the equation and look at the i-th row"

2.

$$\begin{aligned} Cov(X,Y) &= \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)'] \\ &= \mathbf{E}[XY' - X(\mathbf{E}Y)' - \mathbf{E}XY' + \mathbf{E}X(\mathbf{E}Y)'] \\ &= \mathbf{E}(XY') - \mathbf{E}[X(\mathbf{E}Y)'] - \mathbf{E}[\mathbf{E}(X)Y'] + \mathbf{E}[\mathbf{E}X(\mathbf{E}Y)'] \\ &= \mathbf{E}(XY') - \mathbf{E}X(\mathbf{E}Y)' \end{aligned}$$

3.

$$
\begin{aligned}
Var(AX+a) &= Var(AX) + Var(a)^1 \\
&= Var(AX) + 0 \\
&= \mathbf{E}[(AX - A\mathbf{E}X)(AX - A\mathbf{E}X)'] \\
&= \mathbf{E}[AXX'A' - AX(\mathbf{E}X)'A' - A\mathbf{E}XX'A' + A\mathbf{E}X(\mathbf{E}X')A'] \\
&= A\mathbf{E}[XX' - X(\mathbf{E}X)' - \mathbf{E}(X)X' + \mathbf{E}X(\mathbf{E}X')]A' \\
&= A[Var(x)]A'
\end{aligned}
$$

4.

$$
\begin{aligned}
Cov(X,Y) &= \mathbf{E}(XY') - (\mathbf{E}X)(\mathbf{E}Y)' \\
&= \mathbf{E}[(AX + a - \mathbf{E}(AX+a))(BY + b - \mathbf{E}(BY+b))'] \\
&= \mathbf{E}[(AX + a - \underbrace{A\mathbf{E}X - a}_{\text{follows from 1}})(BY + b - B\mathbf{E}(Y) - b)'] \\
&= \mathbf{E}[(AX - A\mathbf{E}(X))(BY - B\mathbf{E}(Y))'] \\
&= A\mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))]B'
\end{aligned}
$$

5. follows from 3

Exercise 1.3

Let $X \in \mathbf{R}^{T\times k}, Y \in \mathbf{R}^{T\times m}, \mathbf{1} = (1,\ldots,1) \in \mathbf{R}^T$. Define

1. $\overline{X} = \frac{1}{T}\mathbf{1}'X$ and $\overline{Y} = \frac{1}{T}\mathbf{1}'Y$
2. $\widehat{Var}(X) = \frac{1}{T}[(X - \mathbf{1}\overline{X})'(X - \mathbf{1}\overline{X})]$
3. $\widehat{Cov}(X,Y) = \frac{1}{T}[(X - \mathbf{1}\overline{X})'(Y - \mathbf{1}\overline{Y})]$

For $A \in \mathbf{R}^{k\times s}, B \in \mathbf{R}^{m\times t}, a \in \mathbf{R}^{1\times s}, b \in \mathbf{R}^{1\times t}$ derive the following transformation rules:

1. $\overline{XA + \mathbf{1}a} = \overline{X}A + a$

   Proof:

$$\begin{aligned}
\overline{XA + \mathbf{1}a} &= \frac{1}{T}[\mathbf{1}'(XA + \mathbf{1}a)] \\
\frac{1}{T}[\mathbf{1}'XA + \underbrace{\mathbf{1}'\mathbf{1}}_{\text{equals } T} a)] &= \overline{X}A + T\frac{1}{T}a \\
&= \overline{X}A + a
\end{aligned}$$

2. $\widehat{Cov}(X,Y) = \frac{1}{T}X'Y - \overline{X}'\overline{Y}$

   Proof:

$$\begin{aligned}
\widehat{Cov}(X,Y) &= \frac{1}{T}[(X - \mathbf{1}\overline{X})'(Y - \mathbf{1}\overline{Y})] \\
&= \frac{1}{T}[X'Y - X'\mathbf{1}\overline{Y} - \overline{X}'\mathbf{1}'Y + \overline{X}'\mathbf{1}'\mathbf{1}\overline{Y}] \\
&= \frac{1}{T}X'Y - \underbrace{\frac{1}{T}X'\mathbf{1}}_{\overline{X}'}\overline{Y} - \overline{X}'\underbrace{\frac{1}{T}\mathbf{1}'Y}_{\overline{Y}} + \frac{1}{T}T\overline{X}'\overline{Y} \\
&= \frac{1}{T}X'Y - 2\overline{X}'\overline{Y} + \overline{X}'\overline{Y} \\
&= \frac{1}{T}X'Y - \overline{X}'\overline{Y}
\end{aligned}$$

3. $\widehat{Cov}(XA + \mathbf{1}a, YB + \mathbf{1}b) = A'\widehat{Cov}(X,Y)B$

   Proof:

$$\begin{aligned}
\widehat{Cov}(XA + \mathbf{1}a, YB + \mathbf{1}b) &= \frac{1}{T}[(XA + \mathbf{1}a - (\mathbf{1}\overline{X}A + \mathbf{1}a))'(YB + \mathbf{1}b - (\mathbf{1}\overline{Y}B + \mathbf{1}b))] \\
&= \frac{1}{T}[(XA - \mathbf{1}\overline{X}A)'(YB - \mathbf{1}\overline{Y}B)] \\
&= A'\frac{1}{T}[(X - (\mathbf{1}\overline{X})'(Y - \mathbf{1}\overline{Y})]B \\
&= A'\widehat{Cov}(X,Y)B
\end{aligned}$$

4. $\widehat{Var}(XA + \mathbf{1}a) = A'\widehat{Var}(X)A$

   Proof:

$$\begin{aligned}
\widehat{Var}(XA + \mathbf{1}a) &= \frac{1}{T}[XA + \mathbf{1}a - \mathbf{1}\bar{(}XA + \mathbf{1}a)]'[XA + \mathbf{1}a - \mathbf{1}(overlineXA + \mathbf{1}a)] \\
&= \frac{1}{T}[XA + \mathbf{1}a - \mathbf{1}(\overline{X}
\end{aligned}$$

Exercise 1.4

We start with a singular value decomposition (SVD) of $X \in \mathcal{R}^{T\times k}$, ie. we have $U \in \mathcal{R}^{T\times T}$ $V \in \mathcal{R}^{k\times k}$, both orthogonal matrices, and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$, where $\sigma_i = \sqrt{\lambda_i}$ with λ_i's being the Eigenvalues of $X'X$ such that

$$X = U\Sigma V' \,.$$

We have to show $X'X\beta = X'y$. We plug in the SVD of X and get

$$\begin{aligned}
X'X\beta &= X'y \\
V\Sigma'\Sigma V'\beta &= V\Sigma' U'y \qquad |V'\cdot \\
\Sigma'\Sigma(V'\beta) &= \Sigma' U'y \\
\begin{pmatrix} \lambda_1 (V'\beta)_1 \\ \dots \\ \lambda_r (V'\beta)_r \\ 0 \\ \dots \\ 0 \end{pmatrix} &= \begin{pmatrix} \sigma_1 (U'y)_1 \\ \dots \\ \sigma_r (U'y)_r \\ 0 \\ \dots \\ 0 \end{pmatrix}
\end{aligned}$$

Define $\Sigma^+ := \text{diag}(1/\sigma_1, \dots, 1/\sigma_r, 0, \dots, 0)$. By this we know that $\beta := V\Sigma^+ U'y$ solves the normal equations.
(Note: $\Sigma^+\Sigma^+\Sigma = \Sigma^+$)

Exercise 1.5

Show that $X(X'X)^{-1}X'$ is the orthogonal projector on the column space spanned by X and show that $I - X(X'X)^{-1}X'$ is the projector on the ortho-complement of the column space spanned by X.

Assumption: $(X'X)^{-1}$ is invertible. Define $P_1 := X(X'X)^{-1}X'$ and $P_2 = X(X'X)^{-1}\underbrace{(X'X)(X'X)^{-1}}_{I}X' = P_1$.

Proof:

$$< a, Pb >= a'X(X'X)^{-1}X'b = X[(X'X)^{-1}]'X'a =< Pa, b > \tag{2}$$

So P is symmetric.

Secondly, we show that P projects on the space Xb by showing that the remainder term $a - Pa$ is orthogonal to the space Xb.

$$< a - Pa, Xb >= (a - Pa)'Xb \tag{3}$$

$$= a'Xb - a'X \underbrace{(X'X)^{-1}X'X}_{I} b \tag{4}$$

$$= a'Xb - a'Xb = 0 \tag{5}$$

$(I - P)a = a - Pa$ $(I - P)^2 = I - 2P + P^2 = I - P$ P is symmetric implies that I-P is symmetric.
Showing that $(I - P)a$ for some given a projects on the orthocomplement of Xb is equivalent to showing that $(I - P)a$ is orthogonal to Xb which is algebraically the same as has been demonstrated above.

Exercise 1.6

Part (i)

Suppose β^+ is not a solution to the normal equation. Then:

$$X'X\beta^+ \neq X'y \quad \text{where } \beta^+ = (X'X)^+X'y$$

This implies:

$$(X'X)(X'X)^+X'y \neq X'y$$

By using the singular value decomposition from exercise 1.4 we know $X = U\Sigma V'$ where V is the eigenvector matrix

$$O\Lambda O'O\Lambda^+O'O\Sigma'U'y \neq O\Sigma'U'y$$

$$OI_r\Sigma'U'y \neq O\Sigma'U'y$$

$$X'y \neq X'y$$

which is a contradiction.

Part (ii)

Show that a given β with $X'X\beta = 0$ implies $\beta'\beta^+ = 0$:

For this we show that $X'X\beta = 0$ implies $X'X\beta^+ = 0$:

$$X'X\beta = O\Lambda O'\beta = \mathbf{0}$$

$$O'O\Lambda O'\beta = \Lambda O'\beta = O'\mathbf{0} = \mathbf{0}$$

since O is orthonormal. Furthermore we know that $\Lambda^+ = \Lambda^+\Lambda^+\Lambda$, so:

$$\Lambda^+O'\beta = \Lambda^+\Lambda^+\mathbf{0} = \mathbf{0}$$

$$O\Lambda^+O'\beta = (X'X)^+\beta = \mathbf{0}$$

the transpose of the latter term is equally zero: $\beta'(X'X)^+ = \mathbf{0}'$. So we have

$$\beta'\beta^+ = \beta'(X'X)^+X'y = \mathbf{0'X'y} = \mathbf{0}$$

Part (iii)

Show that $||\beta^+|| \leq ||\beta||$ where β is a solution to $X'X\beta = X'y$:

$$||\beta^+|| = ||(X'X)^+X'y|| = ||(X'X)^+(X'X)\beta|| =$$

$$= ||O\Lambda^+O'O\Lambda O'|| = ||O\Lambda^+\Lambda O'||$$

since $(X'X)^+ = O\Lambda^+O'$ and $(X'X) = O\Lambda O'$. Moreover $O'O = I$ since O is orthonormal.

Denote with I_r the "pseudo-identity matrix" as the matrix with the first r entries in thediagonal equal to one, and the reamining entries equal to zero.

$$I_r = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots \\ 0 & 1 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & \dots & 1 & 0 & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & & & & \ddots \end{pmatrix}$$

As can easily be seen $\Lambda^+\Lambda = I_r$, and $OI_rO' = I_r$. So:

$$||O\Lambda^+\Lambda O'|| = ||I_r\beta|| \ \leq \ ||\beta||$$

So $||\beta^+|| \leq ||\beta||$. The latter conclusion follows from the fact that $I_r\beta$ is a vector with only the first r entries equal to those of β while the entries from $r+1$ to k are zero.

Exercise 1.7

(1) Show that R^2 as defined in class for inhomogenous regression (including the constant term) is equal to

$$R^2 = r^2_{y\widehat{y}} = \frac{s^2_{y\widehat{y}}}{s_{yy}s_{\widehat{y}\widehat{y}}}$$

Definitions:

1. $s_{y\widehat{y}} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \overline{y})(\widehat{y}_i - \overline{\widehat{y}})$
2. $s_{yy} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \overline{y})^2$
3. $R^2 = \frac{s_{\widehat{y}\widehat{y}}}{s_{yy}}$

Hint: Show that $s_{y\widehat{y}} = s_{\widehat{y}\widehat{y}}$

Starting with the definitions We know that $T\overline{Y}^2 = T\overline{\widehat{Y}}^2$ so $\overline{\widehat{Y}} = \overline{Y}$. From

$$s_{Y\widehat{Y}} = (Y'\widehat{Y} - T\overline{Y}\overline{\widehat{Y}}) = (Y'\widehat{Y} - T\overline{Y}^2)$$

$$s_{\widehat{Y}'\widehat{Y}} = (\widehat{Y}'\widehat{Y} - T\overline{\widehat{Y}}^2) = (\widehat{Y}'\widehat{Y} - T\overline{Y}^2)$$

To show $< Y, \widehat{Y} > = < \widehat{Y}, \widehat{Y} >$.

Proof:

$$
\begin{aligned}
< \widehat{Y}, \widehat{Y} > &= < Y - \widehat{u}, \widehat{Y} > \\
&= < Y, \widehat{Y} > - \underbrace{< \widehat{u}, \widehat{Y} >}_{0} \\
&= < Y, \widehat{Y} >
\end{aligned}
$$

(2) Show that $R^2 = 0$ if the constant ist the only regressor

Proof:

If the constant is the only regressor then $X \in \mathbf{R}^{T\times 1}$ so the linear regression model looks like

$$
Y_{T\times 1} = X_{T\times 1}\beta_{T\times 1} + u_{T\times 1}
$$

So X is a column vector of dimension $T \times 1$ with $x_i = 1 \ \forall i = 1, \ldots, T$ which we will denote as $\mathbf{1}$. The least square estimator $\beta_{LS} = (X'X)^{-1}X'Y$ will in this case look like:

$$
\begin{aligned}
\beta_{LS} &= [\mathbf{1}'\mathbf{1}]^{-1}\mathbf{1}'Y \\
&= (T)^{-1}\mathbf{1}'Y \\
&= [\frac{1}{T}, \frac{1}{T}, \frac{1}{T}, \ldots, \frac{1}{T},][Y_1, Y_2, Y_3, \ldots, Y_T]' \\
&= \frac{1}{T}\sum_{i=1}^{T} Y_i \\
&= \overline{Y}
\end{aligned}
$$

Hence $\widehat{Y} = X\beta_{LS} = \mathbf{1}\overline{Y} = [\overline{Y}, \overline{Y}, \ldots, \overline{Y}]'$. If we reconsider the expression for the $\mathcal{R}^2$, then it will be zero if $s_{Y\widehat{Y}} = 0$.

Calculating $s_{Y\widehat{Y}}$ for our specific $\overline{Y}$ gives us:

$$\begin{aligned} s_{Y\widehat{Y}} &= \frac{1}{T}\sum_{i=1}^{T}(Y_i - \overline{Y})(\widehat{Y}_i - \overline{\widehat{Y}}) \\ &= \frac{1}{T}\sum_{i=1}^{T}(Y_i - \overline{Y})(\overline{Y} - \overline{Y}) \\ &= \frac{1}{T}\sum_{i=1}^{T}(Y_i - \overline{Y}) \times 0 \\ &= 0 \end{aligned}$$

So R^2 will always be zero if we regress Y on simply the constant.

Exercise 1.8

We have to show

$$(y - X\beta)'(y - X\beta) = (y - X\widehat{\beta})'(y - X\widehat{\beta}) + (\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})$$

Multiplying $(y - X\beta)'(y - X\beta)$ and exanding with $X\widehat{\beta}$ yields

$$
\begin{aligned}
(y-X\beta)'(y-X\beta) &= [(y-X\widehat{\beta})-(X\beta-X\widehat{\beta})]'[(y-X\widehat{\beta})-(X\beta-X\widehat{\beta})] \\
&= [(y-X\widehat{\beta})'-(X(\beta-\widehat{\beta}))'][(y-X\widehat{\beta})-(X(\beta-\widehat{\beta}))] \\
&= (y-X\widehat{\beta})'(y-X\widehat{\beta})-\underbrace{(y-X\widehat{\beta})'}_{y-\widehat{y}}(X(\beta-\widehat{\beta}))-(X(\beta-\widehat{\beta}))'\underbrace{(y-X\widehat{\beta})}_{y-\widehat{y}} \\
&+(X(\beta-\widehat{\beta}))'(X(\beta-\widehat{\beta})) \\
&= [(y-X\widehat{\beta})'(y-X\widehat{\beta})+((\beta-\widehat{\beta})'X') \\
&(X(\beta-\widehat{\beta}))]-(y-\widehat{y})'(X\beta-\widehat{y})-(X\beta-\widehat{y})'(y-\widehat{y}) \\
&= [(y-X\widehat{\beta})'(y-X\widehat{\beta}) \\
&+((\beta-\widehat{\beta})'X')(X(\beta-\widehat{\beta}))]-\widehat{u}'(X\beta-\widehat{y})-(X\beta-\widehat{y})'\widehat{u} \\
&= [(y-X\widehat{\beta})'(y-X\widehat{\beta}) \\
&+((\beta-\widehat{\beta})'X')(X(\beta-\widehat{\beta}))]-\underbrace{\widehat{u}'X\beta}_{0}+\underbrace{\widehat{u}'\widehat{y}}_{0} \\
&-\underbrace{(X\beta)'\widehat{u}}_{\beta'X'\widehat{u}=0}+\underbrace{\widehat{y}'\widehat{u}}_{0} \\
&= [(y-X\widehat{\beta})'(y-X\widehat{\beta})+((\beta-\widehat{\beta})'X')(X(\beta-\widehat{\beta}))]
\end{aligned}
$$

Exercise 1.9

Show the second claim of item (iii) of the Frisch-Waugh theorem as discussed.

Frisch Waugh Theorem: We partition our regressor matrix X into $X = [X_1, X_2]$ with $X_1 \in \mathbf{R}^{T\times k_1}$, $X_2 \in \mathbf{R}^{T\times k_2}$ and are assuming that $rk(X) = k_1 + k_2$. Then the residuals of

1. $y = X_1\beta_1 + X_2\beta_2 + u$

   are the same as when regressing

2. $\widehat{\beta}_2 = (\widetilde{X_2}'\widetilde{X_2})^{-1}(\widetilde{X_2}'\widetilde{y})$

with $P_1 = (X_1'X_1)^{-1}X_1'$ and $M_1 = I - P_1$. Here we regress first y on X_1 and denote the residuals of this regression as $\widetilde{y} = M_1 y$. In a second step we then regress X_2 on X_1 and again compute the residuals denoted as $\widetilde{X} = M_1 X_2$. In a third step we use the formerly computed residuals and run the regression as stated in (2). In the lecture it was shown that the residuals of (2) are the same as the ones of (1). We are now asked to show that when running

3. $\widehat{\beta}_2 = (\widetilde{X_2}'\widetilde{X_2})^{-1}(\widetilde{X_2}'y)$

the residuals of (3) are not equal with that of (1)=(2). In (3) we use the original y-variable instead of $\widetilde{y}$.

Proof:

Write Normal Equations in a partitioned form:

$$(1^*)\ X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y$$

$$(2^*)\ X_1\beta_1 + X_2'X_2\beta_2 = X_2'y$$

Now consider the first equation (1*):

$$X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y \tag{6}$$

$$X_1\beta_1 + P_1X_2\beta_2 = P_1y \tag{7}$$

$$X_1\beta_1 = -P_1X_2\beta_2 + P_1y \tag{8}$$

To get from (1) to (2) we have to premultiply (1) by $X_1(X_1'X_1)^{-1}$. Now look at equation (2*) and plug in the expression for $X_1\beta_1$:

$$X_1\beta_1 + X_2'X_2\beta_2 = X_2'y \tag{9}$$

$$X_2'[-P_1X_2\beta_2 + P_1y] + X_2'X_2'X_2\beta_2 = X_2'y \tag{10}$$

$$-X_2'P_1X_2\beta_2 + X_2'P_1y + X_2'X_2\beta_2 = X_2'y \tag{11}$$

$$-X_2'P_1X_2\beta_2 + X_2'X_2\beta_2 = X_2'y - X_2'P_1y \tag{12}$$

$$X_2' \overbrace{[I-P_1]}^{\text{projector is idempotent and symmetric}} X_2\beta_2 = X_2'[I-P_1]y \tag{13}$$

$$X_2'[I-P_1]'[I-P_1]X_2\beta_2 = X_2'[I-P_1]y \tag{14}$$

$$\widetilde{X_2}'\widetilde{X_2}\beta_2 = \widetilde{X_2}'y \tag{15}$$

$$\widehat{\beta}_2 = (\widetilde{X_2}'\widetilde{X_2})^{-1}(\widetilde{X_2}'y) \tag{16}$$

The residuals $\widetilde{y} = \widehat{u} = \widetilde{y} - \widetilde{X_2}\widehat{\beta}_2$ do not equal $u(*) = y - \widetilde{X_2}\widehat{\beta}_2$.Only in the case when y equals $\widetilde{y}$.

Exercise 1.10

$$\begin{aligned}
CC' &= (LX^+)(LX^+)' + (C - LX^+)(C - LX^+)' \\
&= (LX^+)(LX^+)' + CC' + (LX^+)(LX^+)' - C(LX^+)' - (LX^+)C' \\
&= CC' + 2L(X'X)^{-1}X'\left[(X'X)^{-1}X'\right]' L' - CX(X'X)^{-1}L' - L(X'X)^{-1}X'C' \\
&= CC' + 2L(X'X)^{-1}X'X(X'X)^{-1}L' - CX(X'X)^{-1}L' - L(X'X)^{-1}X'C' \\
&= CC' + 2L(X'X)^{-1}L' - L(X'X)^{-1}L' - L(X'X)^{-1}L' \\
&= CC'
\end{aligned}$$

Using the following facts:

- as X has full rank, we know that $X^+ = (X'X)^{-1}X'$,
- $\left[(X'X)^{-1}\right]' = \left[(X'X)'\right]^{-1} = \left[X'X\right]^{-1}$,

- $CX = L$ and therefore $X^{'}C^{'} = L^{'}$.

Exercise 1.11

The mean squared error (MSE) of an estimator $\tilde{\beta}$ of β is defined as $MSE(\tilde{\beta}) = \mathbf{E}[(\tilde{\beta}-\beta)'(\tilde{\beta}-\beta)]$. Show the following claims:

(i) If $\tilde{\beta}$ is unbiased with VCV $\Sigma_{\tilde{\beta}\tilde{\beta}}$, then it holds that $MSE(\tilde{\beta}) = tr(\Sigma_{\tilde{\beta}\tilde{\beta}})$, where tr denotes the trace of a matrix.

For any estimator of $\beta \in \mathbf{R}^k$ we can rewrite

$$MSE(\tilde{\beta}) = \mathbf{E}[(\tilde{\beta}-\beta)'(\tilde{\beta}-\beta)] = \mathbf{E}[\sum_{i=1}^{k}(\tilde{\beta}_i-\beta_i)^2]$$

Now since $\tilde{\beta}$ is unbiased, $\mathbf{E}(\tilde{\beta}) = \beta$, and thus $\Sigma_{\tilde{\beta}\tilde{\beta}} = \mathbf{E}[(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)']$.

Further $tr(\Sigma_{\tilde{\beta}\tilde{\beta}}) = \sum_{i=1}^{k}[\mathbf{E}(\tilde{\beta}_i-\beta_i)^2] = \mathbf{E}[\sum_{i=1}^{k}(\tilde{\beta}_i-\beta_i)^2] = MSE(\tilde{\beta})$ QED.

(ii) Let $\tilde{\beta}_1$ and $\tilde{\beta}_2$ be two unbiased estimators with covariance matrices $\Sigma_{\tilde{\beta}_1\tilde{\beta}_1}$ and $\Sigma_{\tilde{\beta}_2\tilde{\beta}_2}$. Show that it holds that

$$\Sigma_{\tilde{\beta}_1\tilde{\beta}_1} \leq \Sigma_{\tilde{\beta}_2\tilde{\beta}_2} \;\Rightarrow\; MSE(\tilde{\beta}_1) \leq MSE(\tilde{\beta}_2)$$

What does this imply for the OLS estimator?

Define $\Delta := \Sigma_{\tilde{\beta}_2\tilde{\beta}_2} - \Sigma_{\tilde{\beta}_1\tilde{\beta}_1}$. Since $\Sigma_{\tilde{\beta}_1\tilde{\beta}_1} \leq \Sigma_{\tilde{\beta}_2\tilde{\beta}_2}$, Δ must be non-negative definite.

Now, using the results from (i), write

$$MSE(\tilde{\beta}_2) - MSE(\tilde{\beta}_1) = tr(\Sigma_{\tilde{\beta}_2\tilde{\beta}_2}) - tr(\Sigma_{\tilde{\beta}_1\tilde{\beta}_1}) =$$

$$tr(\Sigma_{\tilde{\beta}_2\tilde{\beta}_2} - \Sigma_{\tilde{\beta}_1\tilde{\beta}_1}) = tr(\Delta) = \sum_{i=1}^{k}(e_i'\Delta e_i) \geq 0$$

where $e_i \in \mathbf{R}^k$ is a vector with 1 in the i-th row and all zeros else. Because Δ is non-negative definite, all the elements in this sum must be non-negative, and therefore also the total sum. Thus $MSE(\tilde{\beta}_1) \leq MSE(\tilde{\beta}_2)$. QED

Since the OLS $\hat{\beta}$ has the "smallest" VCV matrix among all linear unbiased estimators of β, this implies that it also has smaller or equal MSE among this class of estimators.

(iii) Minimize $MSE(\tilde{\beta})$ over all linear unbiased estimatiors $\tilde{\beta}$. From the lecture we know that $\Sigma_{\tilde{\beta}\tilde{\beta}} = \sigma^2 DD'$ for all unbiased estimators of β, $\tilde{\beta} = Dy$ $(DX = I)$.
From (i):

$$MSE(\tilde{\beta}) = tr(\Sigma_{\tilde{\beta}\tilde{\beta}}) = tr(\sigma^2 DD')$$

Using the decomposition lemma, we can write

$$DD' = (X^+)(X^+)' + (D - X^+)(D - X^+)'$$

hence

$$tr(\sigma^2 DD') = tr\left(\sigma^2(X^+)(X^+)' + \sigma^2(D - X^+)(D - X^+)\right) =$$

$$= tr\left(\sigma^2(X^+)(X^+)'\right) + tr\left(\sigma^2(D - X^+)(D - X^+)\right) =$$

$$= \sigma^2 \left[tr \left(\underbrace{(X^+)(X^+}_{\text{independent of D and pos. semi-def.}})' \right) + tr \left(\underbrace{(D - X^+)}_{R} \underbrace{(D - X^+)}_{R'} \right) \right]$$

We minimize this expression over D, where R is $R = (r_1', \ldots, r_T')$.

$$tr(RR') = \sum_{i=1}^{T} ||r_i||^2 \geq 0 \text{ since } ||r_i||^2 \geq 0 \ \forall i$$

Now $tr(RR') = 0$ which is equivalent to $r_i' = (0, \ldots, 0)$ $\forall i$ which is equivalent to $D = X^+$.

This implies that $tr(\sigma^2 DD')$ is minimized for $D = X^+$.