Stat 230 HW3. Yifan Zheng 3035272865

1. Lec 6. Q1.

1 Another (simpler) proof of the FWL Theorem From the OLS decomposition of the long regression $\hat{Y} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon}$, first show that

$$(1) \quad (I_n - H_1)Y = (I_n - H_1)X_2\hat{\beta}_2 + \hat{\varepsilon},$$

then show that

$$(2) \quad X_2^{\mathrm{T}}(I_n - H_1)Y = X_2^{\mathrm{T}}(I_n - H_1)X_2\hat{\beta}_2.$$

(3) The FWL Theorem then follows immediately. The FWL Theorem implicitly uses the fact that $X_2^{\mathrm{T}}(I_n - H_1)X_2$ is invertible. Why is it true?

(1).
$$
\begin{aligned}
(I_n - H_1)Y &= (I_n - H_1)(X\hat{\beta} + \hat{\varepsilon}) \\
&= (I_n - H_1)(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon}) \quad \downarrow (\text{because } (I_n - H_1)X_1\hat{\beta}_1 = 0) \\
&= (I_n - H_1)X_2\hat{\beta}_2 + (I_n - H_1)\hat{\varepsilon} \quad \downarrow (\text{because } \hat{\varepsilon} \perp C(X).) \\
&= (I_n - H_1)X_2\hat{\beta}_2 + \hat{\varepsilon}
\end{aligned}
$$

(2).
$$
\begin{aligned}
X_2^T(I_n - H_1)Y &= X_2^T(I_n - H_1)\cdot X\hat{\beta} + X_2^T\cdot\hat{\varepsilon} \quad (\text{because } (1)). \\
&= X_2^T(I_n - H_1)\cdot(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) \quad \downarrow (\text{because } (I_n - H_1)X_1\hat{\beta}_1 = 0. \\
&= X_2^T(I_n - H_1)\cdot X_2\hat{\beta}_2
\end{aligned}
$$

(3) Suppose $X_2^T(I_n - H_1)X_2$ is invertible

$$
\begin{aligned}
\hat{\beta}_2 &= (X_2^T(I_n - H_1)X_2)^{-1} X_2^T(I_n - H_1)Y. \quad (\text{because } (2)) \\
&= (\tilde{X}_2^T\tilde{X}_2)^{-1}\tilde{X}_2^T Y \quad (\text{where } \tilde{X}_2 = (I_n - H_1)X_2) \\
&= (\tilde{X}_2^T\tilde{X}_2)^{-1}\tilde{X}_2^T\tilde{Y} \quad (\text{where } \tilde{Y} = (I_n - H_1)Y)
\end{aligned}
$$

2. Lec 6. Q3.

3 Multivariate regression via univariate regressions The FWL Theorem states that the OLS coefficient in the long regression can be obtained from several short regressions. Consider the most extreme case, if you only know how to compute univariate regressions, how can you compute $\hat\beta_j$, the j-th coordinate in the long regression?

Hint: The coefficient in the OLS fit of a vector $\overset{Y}{a}$ on a vector $\overset{X}{b}$ equals $a^\top b/b^\top b$. $\frac{Y^TX}{X^TX}$

Suppose $X=(X_1,\cdots,X_p)$. $\beta=(\beta_1\cdots\beta_p)^T$

(1) If $X_1\cdots X_p$ are orthogonal, then $\hat\beta_j=\dfrac{Y^TX_j}{X_j^TX_j}$

(2). If they are not orthogonal.

Let $W_1=X_1$,

$$W_2=X_2-\frac{X_2^TX_1}{X_1^TX_1}X_1$$

$$W_3=X_3-\frac{X_3^TX_2}{X_2^TX_2}\cdot X_2-\frac{X_3^TX_1}{X_1^TX_1}X_1$$

$$\vdots$$

$$W_p=X_p-\frac{X_p^TX_{p-1}}{X_{p-1}^TX_{p-1}}X_{p-1}-\cdots-\frac{X_p^TX_1}{X_1^TX_1}X_1$$

Then $W_1,\cdots,W_p$ are orthoganol, and $W_1\cdots W_p\in C(X)$.

$$\hat\beta_j=\frac{Y^TW_j}{W_j^TW_j}$$

3. Prove Theorem 1.

Theorem 1. For $Y, X, W \in \mathbb{R}^n$,

$$\hat{\rho}_{yx|w} = \frac{\hat{\rho}_{yx} - \hat{\rho}_{yw}\hat{\rho}_{xw}}{\sqrt{1-\hat{\rho}_{yw}^2}\sqrt{1-\hat{\rho}_{xw}^2}}.$$

- Run OLS of X on $(1, W)$ $\quad \hat{X} = \bar{X} + \rho_{xw}\cdot\frac{\hat{\sigma}_x}{\hat{\sigma}_w}(W-\bar{W})$.

$$\Rightarrow \hat{\varepsilon}_x = X - \bar{X} - \rho_{xw}\cdot\frac{\hat{\sigma}_x}{\hat{\sigma}_w}(W-\bar{W}) = C_n X - \frac{X^T C_n W}{W^T C_n W}\cdot C_n W.$$

- Run OLS of Y on $(1, W)$. $\quad \hat{Y} = \bar{Y} + \rho_{Yw}\cdot\frac{\hat{\sigma}_Y}{\hat{\sigma}_w}(W-\bar{W})$

$$\Rightarrow \hat{\varepsilon}_Y = Y - \bar{Y} - \rho_{Yw}\cdot\frac{\hat{\sigma}_Y}{\hat{\sigma}_w}(W-\bar{W}) = C_n Y - \frac{Y^T C_n W}{W^T C_n W}\cdot C_n W$$

- Hence, $\hat{\varepsilon}_x^T \hat{\varepsilon}_Y = \left[C_n X - \frac{X^T C_n W}{W^T C_n W}\cdot C_n W\right]^T \cdot \left[C_n Y - \frac{Y^T C_n W}{W^T C_n W}\cdot C_n W\right]$

$$= X^T C_n Y - \frac{X^T C_n W}{W^T C_n W}\cdot W^T C_n Y - \frac{Y^T C_n W}{W^T C_n W}\cdot X^T C_n W + \frac{X^T C_n W \cdot Y^T C_n W}{W^T C_n W}$$

$$= X^T C_n Y - \frac{X^T C_n W \cdot W^T C_n Y}{W^T C_n W}$$

$$= \left[\hat{\rho}_{yx} - \hat{\rho}_{yw}\cdot\hat{\rho}_{xw}\right]\cdot\sqrt{X^T C_n X}\cdot\sqrt{Y^T C_n Y}$$

- And $\|\hat{\varepsilon}_x\| = \sqrt{\hat{\varepsilon}_x^T \varepsilon_x} = \sqrt{\left[C_n X - \frac{X^T C_n W}{W^T C_n W}\cdot C_n W\right]^T\cdot\left[C_n X - \frac{X^T C_n W}{W^T C_n W}\cdot C_n W\right]}$

$$= \sqrt{X^T C_n X - \frac{X^T C_n W}{W^T C_n W}} = \sqrt{X^T C_n X}\cdot\sqrt{1-\hat{\rho}_{xw}}$$

- In the same way, $\|\hat{\varepsilon}_Y\| = \sqrt{Y^T C_n Y}\cdot\sqrt{1-\hat{\rho}_{yw}}$

- Finally, $\hat{\rho}_{yx|w} = \frac{\hat{\varepsilon}_x^T \hat{\varepsilon}_Y}{\|\hat{\varepsilon}_x\|\cdot\|\hat{\varepsilon}_Y\|} = \frac{\hat{\rho}_{Yx} - \hat{\rho}_{xw}\cdot\hat{\rho}_{Yw}}{\sqrt{1-\hat{\rho}_{xw}}\cdot\sqrt{1-\hat{\rho}_{yw}}}$