Problem set will be assigned today:
To obtain .tex files, e-mail Katerina

$Y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon$

$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$

$Var(\hat{\beta}_1 | X_1, X_2) = (X_1' M_2 X_1)^{-1} X_1' M_2 Var(Y | X_1, X_2) M_2' X_1 (X_1' M_2 X_1)^{-1}$

$= \sigma^2 (X_1' M_2 X_1)^{-1}$

$Var(\hat{\beta}_1 | X_1, X_2) \uparrow$ as $\sigma^2 \uparrow$

$\uparrow$ as $\underbrace{X_1' M_2 X_1}_{\text{sum of squared residuals from regression of } X_1 \text{ on } X_2} \downarrow$

How do we estimate this? We do not know σ^2.

$\sigma^2 = E[\varepsilon_i^2]$

a reasonable estimator is $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$, but we do not observe ε_i's. Thus, we can use $\hat{\tilde{\sigma}}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2$

$\hat{\tilde{\sigma}}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n} \hat{\varepsilon}' \hat{\varepsilon}$

Recall: $\hat{\varepsilon} = MY = M_X (X\beta + \varepsilon) = \underbrace{M_X X}_{=0} \beta + M_X \varepsilon = M_X \varepsilon$

$E[\hat{\varepsilon}' \hat{\varepsilon} | X] = E[\varepsilon' M' M \varepsilon | X] = E[\varepsilon' M \varepsilon | X] = E[tr(\varepsilon' M \varepsilon) | X]$

$= E[tr(M \varepsilon \varepsilon') | X] = tr(E[M \varepsilon \varepsilon' | X]) = tr(M E[\varepsilon \varepsilon' | X])$

$= tr(M \sigma^2 I) = tr(\sigma^2 M) = \sigma^2 tr(M) = \sigma^2 tr(I_n - X(X'X)^{-1} X')$

$= \sigma^2 (n - tr(X(X'X)^{-1} X')) = \sigma^2 (n - tr(X'X(X'X)^{-1}))$

$= \sigma^2 (n - tr(I_k)) = \sigma^2 (n-k) \Rightarrow E[\hat{\varepsilon}' \hat{\varepsilon}] = E[\sigma^2 (n-k) | X] = \sigma^2 (n-k)$

$\Rightarrow E[\hat{\tilde{\sigma}}^2] = \frac{1}{n} \sigma^2 (n-k) \neq \sigma^2$

This shows that $\hat{\sigma}^2 = \frac{1}{n-k}\hat{\varepsilon}'\hat{\varepsilon}$ is unbiased.

$\Rightarrow \widehat{Var}(\hat{\beta}|X) = \hat{\sigma}^2(X'X)^{-1}$ is a random variable

$$E[\widehat{var}(\hat{\beta}|X)|X] = E[\hat{\sigma}^2(X'X)^{-1}|X] = (X'X)^{-1}E[\hat{\sigma}^2|X]$$
$$= \sigma^2(X'X)^{-1}.$$

Prediction:

Given X_0, what is our best guess for Y_0?

If $Y \sim f_Y$, then our best "guess" is the $\hat{Y}$ satisfying

$$\underset{\hat{Y}_0}{\arg\min} E[(Y_0 - \hat{Y}_0)^2] = E[Y_0]$$

$$\widehat{E(Y_0)} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\underset{\hat{Y}_0}{\arg\min} E[|Y_0 - \hat{Y}_0|] = med(Y_0)$$

More complicated prediction problem:

$Y, X \sim f_{X,Y}$

$$\underset{h(X_0)}{\arg\min} E[(Y_0 - h(X_0))^2] = \underbrace{E[Y_0|X_0]}_{\text{Best predictor}}$$ (functional analysis problem)

In the OLS world: $E[Y_0|X_0] = X_0\beta$. Usually, we do not know this. We estimate it with $\widehat{E[Y_0|X_0]} = X_0\hat{\beta}$.

$$\Rightarrow \hat{Y}_0 = \widehat{E[Y_0|X_0]} = X_0\hat{\beta}$$

This justifies why this is a good predictor.

$$E[Y_0 - \hat{Y}_0 | X, X_0] = X_0\beta - X_0\beta = 0$$

$\llcorner$ On average, the prediction error is zero

$$Var(Y_0 - \hat{Y}_0|X, X_0) = Var(Y_0|X) + Var(\hat{Y}_0|X) - 2\underbrace{Cov(Y_0, \hat{Y}_0|X)}_{=0 \text{ by independence assumption}}$$
$$= \sigma^2 + \sigma^2 X_0(X'X)^{-1}X_0'$$
$$= \sigma^2(1 + X_0(X'X)^{-1}X_0') > \sigma^2$$

The additional uncertainty is due to the fact that we have to estimate β.

Lecture 4: Classical Normal Linear Regression Model

Assumptions: Same as before and.

$$\underset{n\times 1}{Y} | X \sim N_n(X\beta, \sigma^2 I_n) \iff \begin{cases} Y = X\beta + \varepsilon \\ \varepsilon | X \sim N_n(0, \sigma^2 I_n) \Rightarrow \varepsilon \sim N_n(0, \sigma^2 I_n) \end{cases}$$

This imposes independence between regressors and errors.

$$\hat\beta|_X = (X'X)^{-1}X'Y|_X = AY|X \sim N_k(AX\beta, A\sigma^2 I_n A')$$
$$= N_k(\beta, \sigma^2 AA')$$
$$= N_k(\beta, \sigma^2(X'X)^{-1})$$

Since $\hat\beta$ is linear in Y and $Y|X$ is normal.

This implies that any linear transformation of $\hat\beta$ is normal.

Let Γ be $p\times k$ and fixed

$$\Gamma\hat\beta | X \sim N_p(\Gamma\beta, \sigma^2 \Gamma(X'X)^{-1}\Gamma')$$

For example, let $\Gamma = [0 \ldots \underset{j\text{th}}{1} \ldots 0]$

$$\Rightarrow \Gamma\hat\beta = \hat\beta_j \sim N(\beta_j, [\sigma^2(X'X)^{-1}]_{jj})$$

Normalizing:

$$Q \sim N_p(\mu, \Omega)$$

Since $\Omega' = \Omega$ and Ω is PD, $\exists C : C'C = \Omega^{-1}$

eg $C = \Lambda^{-1/2}H'$ where Λ is diagonal with eits

H' is matrix of eivrs

Call $C = \Omega^{-1/2}$

$\Rightarrow Z = \Omega^{-1/2}(Q-\mu) \sim N_p(0, I_p)$

$\Rightarrow W = Z'Z \sim \chi^2(p)$ since $Z_i \sim N(0,1)\ \forall i=1,\dots,p$

$\Gamma\hat{\beta}|X \sim N_p(\Gamma\beta, \sigma^2\Gamma(X'X)^{-1}\Gamma')$

$Z = (\sigma^2\Gamma(X'X)^{-1}\Gamma')^{-1/2}(\Gamma\hat{\beta}-\Gamma\beta)$

$\Rightarrow Z|X \sim N_p(0, I_p) \Rightarrow Z \sim N(0, I_p)$

$W = Z'Z = (\Gamma\hat{\beta}-\Gamma\beta)'(\sigma^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta}-\Gamma\beta)$

$\Rightarrow W|X \sim \chi^2(\underset{\text{not a fcn of } X}{p}) \Rightarrow W \sim \chi^2(p)$

For $p=1$, $t = [\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma']^{-1/2}(\Gamma\hat{\beta}-\Gamma\beta)$

$\Rightarrow t|X \sim t(n-k) \Rightarrow t \sim t(n-k)$

For $p>1$, $F = \frac{1}{p}(\Gamma\hat{\beta}-\Gamma\beta)'(\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta}-\Gamma\beta)$

$\Rightarrow F|X \sim F(p, n-k) \Rightarrow F \sim F(p, n-k)$

$n-k \to \infty \Rightarrow t(n-k) \to N(0,1)$

For $n-k \to \infty$, $pF(p, n-k) \to \chi^2(p)$