# Review Session for Econometrics Midterm

Tong Zhou

April 30, 2020

Warning: this notes may contain typos and factual errors.

## Contents

# Notations and Theorems

### Stochastic Order

- $X_n = o_p(1)$ if $X_n \xrightarrow{\mathsf{P}} 0$.
- $X_n = O_p(1)$ (*bounded in probability*) if $\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}(|X_n| > M) = 0$.
- $X_n = O_p(a_n)$ means $\frac{X_n}{a_n} = O_p(1)$.
- $O_p(1)o_p(1) = o_p(1)$ ; $O_p(a_n)O_p(b_n) = O_p(a_n b_n)$ ; $O_p(a_n) + O_p(b_n) = O_p(a_n + b_n) = O_p(\max(a_n, b_n))$.
- If $X_n \xrightarrow{\mathsf{d}} X$, then $X_n = O_p(1)$. $\left(\sqrt{n}(\widehat{\beta} - \beta) = O_p(1)\right)$

### Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \dots, X_n$ be IID with $\mathbb{E}[X_i] = \mu$, $S_n = \sum_{i=1}^n X_i$, then

$$\frac{S_n}{n} = \overline{X}_n \xrightarrow{\mathsf{P}} \mu \quad \text{or} \quad \overline{X}_n - \mu \xrightarrow{\mathsf{P}} 0.$$

### Central Limit Theorem (CLT)

Let $X_1, X_2, \dots, X_n$ be IID with $\mathbb{E}[X_i] = \mu$ and $\mathsf{Var}(X_i) = \sigma^2 < \infty$. Then

$$\frac{S_n - n\mu}{\sqrt{n}} = \sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2)$$

### Slutsky Theorem

If $X_n \xrightarrow{\mathsf{d}} X$ and $Y_n \xrightarrow{\mathsf{P}} c$, then

1. $X_n + Y_n \xrightarrow{\mathsf{d}} X + c$.
2. $X_n Y_n \xrightarrow{\mathsf{d}} cX$.
3. $\frac{X_n}{Y_n} \xrightarrow{\mathsf{d}} \frac{X}{c}$, provided $c \neq 0$.

### Continuous Mapping Theorem

If $f$ is a continuous function and $Y_n \xrightarrow{\mathsf{d}} Y$, then

$$f(Y_n) \xrightarrow{\mathsf{d}} f(Y).$$

### Delta Method

If $\sqrt{n}(X_n - \mu) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2)$, and $f$ is differentiable at $\mu$, then

$$\sqrt{n}(f(X_n) - f(\mu)) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2 f'(\mu)^2).$$

# 2016 Q2

By

$$e_i = \boldsymbol{y_i} - \boldsymbol{x}_i'\widehat{\beta} = \varepsilon_i - \boldsymbol{x}_i'(\widehat{\beta} - \beta) \xrightarrow{\mathsf{P}} \varepsilon_i$$

# 2016 Q3

The sample mean $\bar{y} = \mu + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$.

By the LLN:

$$\bar{y} \xrightarrow{\mathsf{P}} \mu$$

By the CLT:

$$\sqrt{n}(\bar{y} - \mu) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2).$$

The alternative estimator

$$\hat{\mu} = \sum_i w_i y_i = \mu + \sum_i w_i \varepsilon_i.$$

By Chebyshev's inequality, for any $\delta > 0$,

$$\begin{aligned} \mathbb{P}\big(|\hat{\mu} - \mu| > \delta\big) &\leq \frac{\mathbb{E}[(\sum_i w_i \varepsilon_i)^2]}{\delta^2} \\ &= \frac{\sigma^2 \sum w_i^2}{\delta^2} \\ &= \frac{\sigma^2 n(n+1)(2n+1)}{6\delta^2 n^2 (n+1)^2/4} \\ &\to 0 \end{aligned}$$

So $\hat{\mu} \xrightarrow{\mathsf{P}} \mu$.

The asymptotic variance is:

$$n \operatorname{\mathsf{Var}}(\hat{\mu}) = n \cdot \frac{2\sigma^2}{3} \cdot \frac{n(n+1)(2n+1)}{n^2(n+1)^2} \to \frac{4}{3}\sigma^2 > \sigma^2$$

that is

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \frac{4}{3}\sigma^2).$$

# 2019 Q6

Find $\widehat{\beta}_L$ and $\widehat{\beta}_U$ s.t.

$$\beta \in [\operatorname{plim} \widehat{\beta}_L, \operatorname{plim} \widehat{\beta}_U].$$

**Proof**

Since there is no intercept in the regression, we obtain:

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

By the WLLN and Slutsky Theorem,

$$\begin{aligned}
\hat{\beta} \xrightarrow{P} & \frac{\mathbb{E}(xy)}{\mathbb{E}(x^2)} \\
&= \frac{\mathbb{E}\left((x^* + \varepsilon)(\beta x^* + \eta)\right)}{\mathbb{E}(x^* + \varepsilon)^2} \\
&= \beta \frac{\mathbb{E}[(x^*)^2]}{\mathbb{E}\left[(x^*)^2\right] + \mathbb{E}(\varepsilon^2)} \\
&= \beta \left( \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2} \right) \\
&\leq \beta.
\end{aligned}$$

Therefore, we obtained a lower bound for $\beta$: $\operatorname{plim} \hat{\beta}$.

Now, we consider a reverse regression of $x$ on $y$ without intercept:

$$x = \gamma y + u$$

where $x$ and $y$ are generated according to the given process.

Similarly, we have

$$\hat{\gamma} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n y_i^2}.$$

By the WLLN,

$$\begin{aligned}\hat{\gamma} \xrightarrow{P} \frac{\mathbb{E}(yx)}{\mathbb{E}(y^2)} \\ &= \frac{\mathbb{E}\left((\beta x^* + \eta)(x^* + \varepsilon)\right)}{\mathbb{E}(\beta x^* + \eta)^2} \\ &= \frac{\beta \mathbb{E}[(x^*)^2]}{\beta^2 \mathbb{E}\left[(x^*)^2\right] + \mathbb{E}(\eta^2)} \\ &= \frac{\beta \sigma^2_{x^*}}{\beta^2 \sigma^2_{x^*} + \sigma^2_\eta} \equiv \gamma.\end{aligned}$$

By the Continuous Mapping Theorem,

$$\begin{aligned}\frac{1}{\hat{\gamma}} \xrightarrow{P} \frac{1}{\gamma} \\ &= \beta + \frac{\sigma^2_\eta}{\beta \sigma^2_{x^*}} \\ &\geq \beta.\end{aligned}$$

So we obtained an upper bound for $\beta$: $\operatorname{plim} \frac{1}{\hat{\gamma}}$.

Therefore, $\beta \in \left[\operatorname{plim} \hat{\beta}, \operatorname{plim} \frac{1}{\hat{\gamma}}\right]$, if $\beta > 0$. If $\beta < 0$, just reverse the upper and lower bounds.

We can routinely compute the bias, variance and mean squared error of $\hat{\beta}$.

You do not need to read this. Here I want to show the point identification can be gained by the quite strong mutual independence condition. For completeness, I put the proof here.

If the mutual independence is maintained, we shall show that the point identification is gained by exploiting the higher moments.

define

$$z = (x - \mathbb{E}[x])(y - \mathbb{E}[y])$$

Then

$$\beta = \frac{\mathsf{Cov}(z, y)}{\mathsf{Cov}(z, x)}$$

**sketch of proof:**

Since

$$y = x^* \beta + \varepsilon = x\beta + \varepsilon - \eta\beta,$$

then
$$\frac{\mathrm{Cov}(z,y)}{\mathrm{Cov}(z,x)} = \beta + \frac{\mathrm{Cov}(z,\varepsilon-\eta\beta)}{\mathrm{Cov}(z,x)} = \beta + \frac{\mathrm{Cov}(z,\varepsilon)}{\mathrm{Cov}(z,x)} - \beta\frac{\mathrm{Cov}(z,\eta)}{\mathrm{Cov}(z,x)}.$$
The first term covariance terms are:

$$\mathrm{Cov}(z,\varepsilon) = \mathrm{Cov}(xy,\varepsilon) - \mathbb{E}[x]\,\mathrm{Cov}(y,\varepsilon).$$
$$\mathrm{Cov}(z,\eta) = \mathrm{Cov}(xy,\eta) - \mathbb{E}[y]\,\mathrm{Cov}(x,\eta).$$

From the followings:

$$\mathrm{Cov}(xy,\varepsilon) = \mathbb{E}[x]\sigma_\varepsilon^2$$
$$\mathrm{Cov}(y,\varepsilon) = \sigma_\varepsilon^2$$
$$\mathrm{Cov}(xy,\eta) = \beta\mathbb{E}[x^*]\sigma_\eta^2$$
$$\mathrm{Cov}(x,\eta) = \sigma_\eta^2,$$

we have $\mathrm{Cov}(z,\varepsilon) = \mathrm{Cov}(z,\eta) = 0$, which implies

$$\beta = \frac{\mathrm{Cov}(z,y)}{\mathrm{Cov}(z,x)}.$$

Also, $\mathrm{Cov}(z,x) \neq 0$ implies
$$\mathbb{E}\left[(x^* - \mathbb{E}[x^*])^3\right] \neq 0.$$

# 2019 Q1 & Q2

## 2019 Q1

### (a)

When either $\beta_2 = 0$ (i.e. $\mathbb{E}(x^2)\mathbb{E}(x^2y) = \mathbb{E}(x^3)\mathbb{E}(xy)$) or $x$ and $x^2$ are orthogonal and non-zero (i.e. $\mathbb{E}(x^3) = 0$ and $\mathbb{E}(x^2)\mathbb{E}(x^4) \neq 0$), $\gamma_1$ and $\beta_1$ coincide.

### (b)

The only condition under which $\gamma_1 = \theta_1$ is that $\theta_2 = 0$ (i.e. $\mathbb{E}(x^2)\mathbb{E}(x^3y) = \mathbb{E}(x^4)\mathbb{E}(xy)$).

Unlike (a) in which the orthogonality condition, $\mathbb{E}(x^3) = 0$, between regressors allows $\gamma_1 = \beta_1$, the orthogonality between $x$ and $x^3$ (i.e. $\mathbb{E}(x^4) = 0$) renders $\gamma_1 = \theta_1$ impossible, since it implies $x = 0$ with probability 1, and then $\mathbb{E}(x^p) = 0$ for any $p > 0$. This pathological condition defies the fact that $\gamma_1 = \frac{\mathbb{E}(xy)}{\mathbb{E}(x^2)}$ is defined.

**Note**: In both $(a)$ and $(b)$, we can also argue the orthogonality condition by the *Frisch-Waugh-Lovell Theorem* that yields the same results.

### 2019 Q2

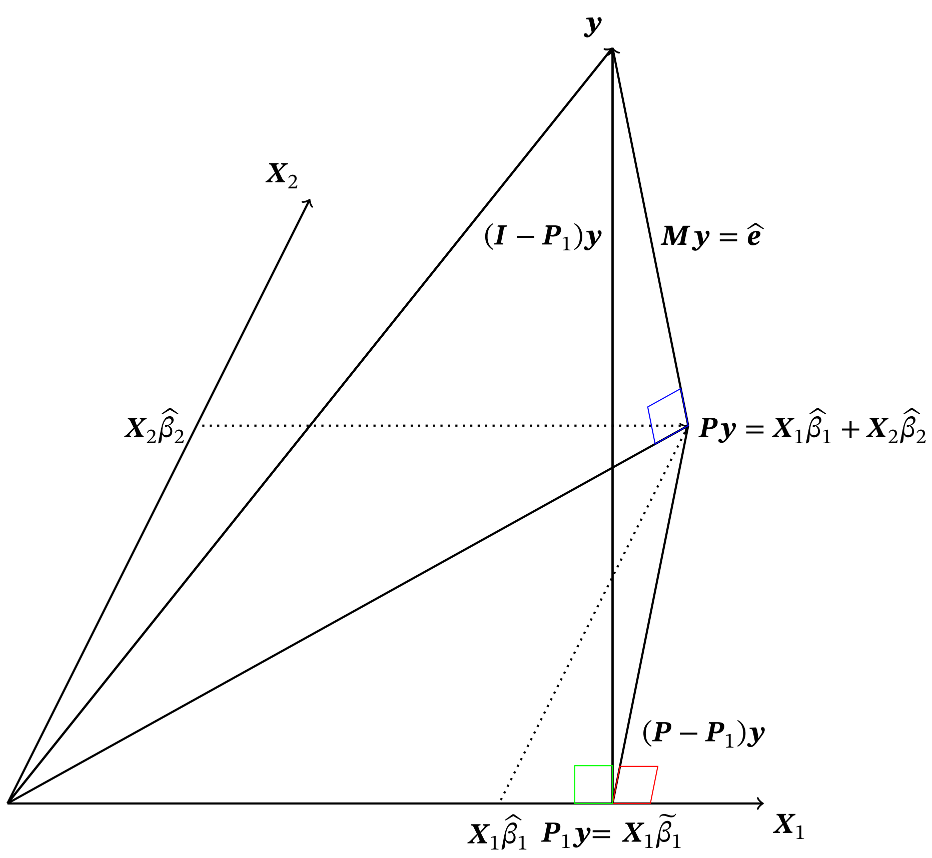
We know:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y} \\
\widehat{\boldsymbol{\beta}}_2 &= (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y} \\
\widetilde{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y} \\
\widetilde{\boldsymbol{\beta}}_2 &= (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}
\end{aligned}
$$

Thus, when $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonal (i.e. $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{X}_2'\boldsymbol{X}_1 = 0$), $\widehat{\boldsymbol{\beta}}_1 = \widetilde{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2 = \widetilde{\boldsymbol{\beta}}_2$.

The other condition is $\boldsymbol{X}_1'\boldsymbol{y} = \boldsymbol{X}_2'\boldsymbol{y} = 0$, under which all estimates are zero.

# Frisch-Waugh-Lovell Theorem



# 2019. Q5 (2017 Q3)

## (a)

By the Delta Method:

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, 4\mu^2\nu^2).$$

## (b)

If $\mu = 0$, $\sqrt{n}\widehat{\beta} \xrightarrow{\mathsf{d}} 0$, or equivalently, $\sqrt{n}\widehat{\beta} \xrightarrow{\mathsf{P}} 0$.

In other words, the normal limit in (a) is degenerate: $\sqrt{n}\widehat{\beta}$ convergens in probability to the constant 0. This is not what we mean by the asymptotic distribution. Thus, we must treat the case $\mu = 0$ separately.

Note that in (a), $\widehat{\beta} - \beta = O_p\left(\frac{1}{\sqrt{n}}\right)$, while in (b) $\widehat{\beta} - \beta = o_p\left(\frac{1}{\sqrt{n}}\right)$.

## (c)

Define random variable $Z \sim \mathcal{N}(0, \nu^2)$, then

$$\sqrt{n}(\widehat{\mu} - \mu) \xrightarrow{\mathsf{d}} Z$$

By the Continuous Mapping Theorem when $\mu = 0$,

$$n\widehat{\beta} \xrightarrow{\mathsf{d}} Z^2$$

since $\frac{Z}{\nu} \sim \mathcal{N}(0, 1)$, then

$$\frac{Z^2}{\nu^2} \sim \chi^2(1)$$

and therefore

$$n\widehat{\beta} \xrightarrow{\mathsf{d}} \nu^2\chi^2(1).$$

It shows that

$$\widehat{\beta} = O_p\left(\frac{1}{n}\right).$$

### (d)

The reason is that when $\mu = 0$, the conventional (first-order) Delta Method breaks down. A finer approximation (i.e. second-order Delta Method) is then in place.

## 2017 Q2 + 2016 Q5

### (1)

By definition, $\hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}$. Then,

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}^2|\boldsymbol{X}] &= \frac{1}{n}\mathbb{E}[\widehat{\boldsymbol{e}}'\hat{\boldsymbol{e}}|\boldsymbol{X}] \\
&= \frac{1}{n}\mathbb{E}[\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}|\boldsymbol{X}] \\
&= \frac{1}{n}\mathbb{E}[\operatorname{tr}\left(\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}\right)|\boldsymbol{X}] \\
&= \frac{1}{n}\mathbb{E}[\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}'\right)|\boldsymbol{X}] \\
&= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\mathbb{E}[\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}]\right) \\
&= \frac{(n-k)}{n}\sigma^2
\end{aligned}
$$

Therefore, $\hat{\sigma}^2$ is biased for $\sigma^2$, and we propose an unbiased estimator:

$$
\frac{n}{(n-k)}\hat{\sigma}^2 = \frac{1}{(n-k)}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \frac{1}{(n-k)}\sum_{i=1}^{n}\hat{e}_i^2.
$$

**(2)**

Since

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_i \hat{e}_i^2 \\
&= \frac{1}{n}\sum_i (y_i - \boldsymbol{x}_i'\widehat{\beta})^2 \\
&= \frac{1}{n}\sum_i \left[e_i - \boldsymbol{x}_i'(\widehat{\beta}-\beta)\right]^2 \\
&= \frac{1}{n}\sum_i e_i^2 - \underbrace{(\widehat{\beta}-\beta)'}_{=o_p(1)}\overbrace{\frac{2}{n}\sum_i(\boldsymbol{x}_i e_i)}^{=o_p(1)} + \underbrace{(\widehat{\beta}-\beta)'}_{=o_p(1)}\overbrace{\frac{1}{n}\sum_i(\boldsymbol{x}_i\boldsymbol{x}_i')}^{=O_p(1)}\underbrace{(\widehat{\beta}-\beta)}_{=o_p(1)} \\
&= \frac{1}{n}\sum_i e_i^2 + o_p(1) \\
&\to \mathbb{E}[e_i^2] = \sigma^2
\end{aligned}
$$

We propose $\hat{\sigma}$. By the Continuous Mapping Theorem, $\hat{\sigma} \xrightarrow{\text{P}} \sigma$.

By Jensen's inequality, however

$$
\mathbb{E}[\hat{\sigma}] \leq \sqrt{\mathbb{E}[\hat{\sigma}^2]} = \sqrt{\frac{(n-k)}{n}}\sigma < \sigma
$$

So $\hat{\sigma}$ is downward biased.

### (3) 2016 Q5

Continuing on (2)

$$
\begin{aligned}
\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_i e_i^2 - \sigma^2\right) \\
& - \underbrace{\sqrt{n}(\widehat{\beta} - \beta)'}_{=O_p(1)} \overbrace{\frac{2}{n}\sum_i(\boldsymbol{x}_i e_i)}^{=o_p(1)} \\
& + \underbrace{\sqrt{n}(\widehat{\beta} - \beta)'}_{=O_p(1)} \overbrace{\frac{1}{n}\sum_i(\boldsymbol{x}_i \boldsymbol{x}_i')}^{=O_p(1)} \underbrace{(\widehat{\beta} - \beta)}_{o_p(1)} \\
& = \sqrt{n}\left(\frac{1}{n}\sum_i e_i^2 - \sigma^2\right) + o_p(1) \\
& \xrightarrow{\mathsf{d}} \mathcal{N}\left(0, \mathsf{Var}(e_i^2)\right).
\end{aligned}
$$

# 2017 Q1

Show that $R_2^2 \geq R_1^2$.

Key: Show $(\boldsymbol{P} - \boldsymbol{P}_1)$ is positive semi-definite.

# 2015 Q1

## (a)

$$
\begin{aligned}
\mathbb{E}[\tilde{\beta}|\boldsymbol{X}] &= \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\mathbb{E}[\boldsymbol{y}|\boldsymbol{X}] \\
&= \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\beta \\
&= \beta
\end{aligned}
$$

## (b)

The variance of GLS estimator $\tilde{\beta}$ is:

$$\begin{aligned}\mathsf{Var}\left(\tilde{\beta}|\boldsymbol{X}\right) &= \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\,\mathsf{Var}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\end{aligned}$$

## (c)

Since

$$\boldsymbol{M}_1\boldsymbol{y} = \underbrace{\boldsymbol{M}_1\boldsymbol{X}}_{=\boldsymbol{0}}\beta + \boldsymbol{M}_1\boldsymbol{e} = \boldsymbol{M}_1\boldsymbol{e},$$

and

$$\begin{aligned}\boldsymbol{M}_1\boldsymbol{y} &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1})\boldsymbol{y} \\ &= \boldsymbol{y} - \boldsymbol{X}\underbrace{(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}}_{=\tilde{\beta}} \\ &= \hat{\boldsymbol{e}},\end{aligned}$$

then $\boldsymbol{M}_1\boldsymbol{e} = \hat{\boldsymbol{e}}$.

## (d)

Plugging $\boldsymbol{M}_1$ of (c) into $\boldsymbol{M}_1'\boldsymbol{\Omega}^{-1}\boldsymbol{M}_1$.

## (e)

$$
\begin{aligned}
\mathbb{E}[s^2|\boldsymbol{X}] &= \frac{1}{n-k}\mathbb{E}[\hat{e}'\Omega^{-1}\hat{e}|\boldsymbol{X}] \\
&= \frac{1}{n-k}\mathbb{E}[\mathrm{tr}(\hat{e}'\Omega^{-1}\hat{e})|\boldsymbol{X}] \\
&= \frac{1}{n-k}\mathbb{E}[\mathrm{tr}(\Omega^{-1}\hat{e}\hat{e}')|\boldsymbol{X}] \\
&= \frac{1}{n-k}\,\mathrm{tr}(\Omega^{-1}M_1\mathbb{E}[ee'|\boldsymbol{X}]M_1') \\
&= \frac{\sigma^2}{n-k}\,\mathrm{tr}(\Omega^{-1}M_1\Omega M_1') \\
&= \frac{\sigma^2}{n-k}\,\mathrm{tr}(\Omega\,\underbrace{M_1'\Omega^{-1}M_1}_{\text{Apply (d)}}) \\
&= \frac{\sigma^2}{n-k}\,\mathrm{tr}(\boldsymbol{I}_n-\boldsymbol{I}_k) \\
&= \sigma^2
\end{aligned}
$$

# 2019 Q4 & 2015 Q2: weighted least square

Before diving into the proof, two assumptions are maintained:

1. $w_i > 0$, for any $i$.
2. $\mathbb{E}[e_i|x_i, w_i] = 0$.

## (1)

$S(\beta)$ can be written as:

$$S(\beta) = (\boldsymbol{y}-\boldsymbol{X}\beta)'\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X}\beta),$$

where $\boldsymbol{W} = \mathrm{diag}(w_1, w_2, \dots, w_n)$.

Then it is easy to show:

$$
\begin{aligned}
\widehat{\beta} &= (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}) \\
&= \left(\sum_{i=1}^{n} w_i\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\sum_{i=1}^{n} w_i\boldsymbol{x}_i y_i\right)
\end{aligned}
$$

### (2)

Under certain regularity conditions,

$$\widehat{\beta} \xrightarrow{\mathsf{P}} \beta$$

### (3)

Under certain regularity conditions,

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{\mathsf{d}} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_\beta)$$

where $\boldsymbol{V}_\beta = \boldsymbol{Q}_{\boldsymbol{xwx}}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}_{\boldsymbol{xwx}}^{-1}$, $\boldsymbol{Q}_{\boldsymbol{xwx}} = \mathbb{E}[w_i \boldsymbol{x}_i' \boldsymbol{x}_i]$, $\boldsymbol{\Omega} = \mathbb{E}[w_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' e_i^2]$.

## 2016 Q1

$\boldsymbol{x}$ should be treated as a constant.

### a.

Let $\hat{\beta} = \boldsymbol{c}'\boldsymbol{y}$. Since

$$\begin{aligned} \mathsf{Var}(\hat{\beta}) &= \boldsymbol{c}' \mathsf{Var}(\boldsymbol{y}) \boldsymbol{c} = \sigma^2 \boldsymbol{c}'\boldsymbol{c} \\ \mathbb{E}[\hat{\beta}] &= \boldsymbol{c}'\boldsymbol{x}\beta, \end{aligned}$$

the mean squared error of $\hat{\beta}$ is

$$\mathrm{MSE}(\hat{\beta}) = \sigma^2 \boldsymbol{c}'\boldsymbol{c} + \beta^2 \left(\boldsymbol{c}'\boldsymbol{x} - 1\right)^2$$

Taking derivative w.r.t. $\boldsymbol{c}$, FOC gives:

$$\sigma^2 \boldsymbol{c} + \beta^2 (\boldsymbol{c}'\boldsymbol{x} - 1)\boldsymbol{x} = \mathbf{0}.$$

Rearrange it:

$$\boldsymbol{c} = \left(\sigma^2 I_n + \beta \boldsymbol{x}\boldsymbol{x}'\right)^{-1} \beta \boldsymbol{x}$$

Apply the **Sherman-Morrison Formula** on page 830 of Hansen's book:

$$\boldsymbol{c} = \frac{\beta^2 \boldsymbol{x}}{\sigma^2 + \beta^2 \boldsymbol{x}'\boldsymbol{x}}$$

Then

$$\hat{\beta} = \frac{\beta^2 \boldsymbol{x}' \boldsymbol{y}}{\sigma^2 + \beta^2 \boldsymbol{x}' \boldsymbol{x}}$$

and

$$\mathrm{MSE}(\hat{\beta}) = \frac{\sigma^2 \boldsymbol{x}' \boldsymbol{x} + \frac{\sigma^4}{\beta^2}}{(\frac{\sigma^2}{\beta^2} + \boldsymbol{x}' \boldsymbol{x})^2}$$

## b

We know the mean squared error of $b$ $\mathrm{MSE}(b) = \sigma^2(\boldsymbol{x}'\boldsymbol{x})^{-1}$.

Their ratio is:

$$\frac{MSE(\hat{\beta})}{MSE(b)} = \frac{\tau^2}{1+\tau^2}, \quad \text{where } \tau^2 = \frac{\beta^2}{\sigma^2/\boldsymbol{x}'\boldsymbol{x}}$$

As $\tau \to \infty$, suppose $\sigma^2$ is fixed, it equivalent to $\beta \to \infty$. As a result,

$$\hat{\beta} \to b \quad \text{or} \quad \mathrm{Bias}(\hat{\beta}) \to 0.$$

The intuition is that as $\beta$ keeps increasing, the bias of $\hat{\beta}$ is going down and therefore the ratio approaches 1.