计量经济学

May 17, 2025

Chapter 1

Lecture 5 Asymptotic Theory for Least Squares

1.1 几种统计概念

1. 依概率：$\lim P[\|Z_n - Z\| \leq \delta] = 1$
2. 依分布：$F_n(u) = p[Z_n \leq u]$
3. 大数定律：$\bar{Y} \to E[Y]$
4. 中心极限定理（CLT）：$\sqrt{n}(\bar{Y} - u) \to N(0, V)$
5. 连续映射定理：$Z_n \to c, g(Z_n) \to g(c)$
6. Delta Method:$\sqrt{n}(g(\hat{u} - g(u))) \to G'\xi, G = \nabla g(u)$

1.2 OSL estimator 与 bets linear projection 的关系

- $\frac{1}{n}\sum_i X_i X_i' \to E[XX']$
- $\frac{1}{n}\sum_i X_i Y_i \to E[XY]$
- 可以说，OLS estimator$\hat{\beta}$ 依概率收敛到 projection coefficientβ。

1.3 Asymptotic Normality

- $\hat{Q}_{XX} = n^{-1}\sum_i X_i X_i' \quad Q_{XX} = E[XX']$
- $\Omega = E[(Xe)(Xe)'] = E[XX'e^2]$
- $\frac{1}{\sqrt{n}}\sum_i X_i e_i \to N(0,\Omega)$
- $\sqrt{n}(\hat{\beta}-\beta) = (\frac{1}{n}\sum_i X_i X_i')^{-1}(\frac{1}{\sqrt{n}}\sum_i X_i e_i) \to Q_{XX}^{-1}N(0,\Omega)$
- $V_\beta = Q_{XX}^{-1}\Omega Q_{XX}^{-1}$，is the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$
- $V_{\hat{\beta}} = \text{var}[\hat{\beta}|X] \quad nV_{\hat{\beta}} \to V_\beta$；$V_{\hat{\beta}}^0 = (X'X)^{-1}\sigma^2$
- 同方差下：$\Omega = E[XX']E[e^2] = Q_{XX}\sigma^2$, 则 $V_\beta = Q_{XX}^{-1}\sigma^2 = V_\beta^0$

1.4 Consistency of Error Variance Estimators

1. $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{e}_i^2 \quad s^2 = (n-k)^{-1}\sum_i \hat{e}_i^2 \quad n^{-1}\sum_i e_i^2 \to \sigma^2$
2. 证明 $\hat{\sigma}^2 \to \sigma^2$，就需利用 $\hat{e}_i = e_i - X_i'(\hat{\beta}-\beta)$ 平方可得。同时也有 $s^2 \to \sigma^2$

1.5 同方差、异方差协方差矩阵估计

1. $\hat{V}_{\hat{\beta}}^0 = (X'X)^{-1}s^2$
2. $\hat{V}_\beta^0 = Q_{XX}^{-1}s^2$ 收敛到 V_β^0
3. $\hat{\Omega} = n^{-1}\sum_i X_i X_i'\hat{e}_i^2 \quad \hat{V}_\beta^{HC0} = \hat{Q}_{XX}^{-1}\hat{\Omega}\hat{Q}_{XX}^{-1}$
4. 同样的有 $n\hat{V}_{\hat{\beta}}^{HC0} = \hat{V}_\beta^{HC0}$

1.6 Functions of Parameters

主要是基于连续映射定理，因此基于 Asymptotic Distribution of Functions of Parameters 往往有：

$$\sqrt{n}(\hat{\theta}-\theta) \to N(0, R'V_\beta R) \quad R = \nabla r(\beta), r(\beta) = \theta$$

注意 T-test 的推导。置信区间不是计量的内容，因此第一轮中跳过

Chapter 2

Lecture 6 Endogeneity

1. structural parameter 就是 β。内生性问题下的估计参数可以说是不收敛的，但不能说是有偏的
2. 经典测量误差：$X = Z + u \quad E[u] = 0$。是有噪声但是无偏测度。
3. 回归中的测量误差：$\beta^* = \beta + \frac{E[Xv]}{E[X^2]}$。前提是 $k = 1$，还可以化为什么形式？什么是 v?
4. 联立方程组误差：建议手动计算推导。

Chapter 3

Lecture 7 Panel Data

面板数据分为平衡面板与非平衡面板。面板数据的列代表同一个体在不同时间，行代码不同个体在同一时间

3.1 混合回归

仔细思考混合回归什么时候是无偏的？当

$$E[e_{it}|X_i] = 0$$

如何推导？写成求和形式而非矩阵形式更加明了。

3.2 随机效应

矩阵形式：

$$Y_i = X_i\beta + 1u_i + \epsilon_i$$

随机效应的六个期望；随机效应模型的方差 $E[e_ie_i'|X_i] = 1_i1_i'\sigma_u^2 + I_i\sigma_\epsilon^2 \sim \Omega_i$，即非对角线元素都为 σ_u^2。该矩阵的维度依赖 T_i 基于上述方差，可以采用 GLS 进行处理（同除 $\Omega^{-\frac{1}{2}}$）所得到最后是同方差。注意,$V_{gls} \leq V_{pool}$（不然为啥要搞广义线性回归）。

3.3 固定效应

随机效应模型中 u_i 常被看成是遗漏变量，如果此变量与 X_i 存在相关性时，估计系数会严重有偏（详见-1 的例子），也即 $E[X_iu_i] \neq 0$，我说的。因此，固定效应模型的作用就是处

理此类情形。它需要的假设要比随机效应模型更弱：

$$E[X_{is}\epsilon_{it}] = 0$$

注意 s 是变化的。如何消除 u_i 呢？如果对两边取均值，可以得到：

$$\bar{Y}_i = \bar{X}_i'\beta + u_i + \bar{\epsilon}_i$$

原因在于 u_i 与不随时间变化。与原回归相减得：

$$\begin{aligned}(Y_{it} - \bar{Y}_i) &= (X_{it} - \bar{X}_i)'\beta + (\epsilon_{it} - \bar{\epsilon}_i) \\ \dot{Y}_{it} &= \dot{X}_{it}'\beta + \dot{\epsilon}_{it}\end{aligned}$$

由

$$\begin{aligned}\dot{Y}_i &= Y_i - 1_i\bar{Y}_i \\ &= Y_i - 1_i(1_i'1_i)^{-1}1_i'Y_i \\ &= M_iY_i\end{aligned}$$

（注意到 $\frac{1}{n} = (1_i'1_i)^{-1}$ 是标量）所以得出：

$$\begin{aligned}\beta_{fe} &= (\sum_i\sum_t \dot{X}_{it}\dot{X}_{it}')^{-1}(\sum_i\sum_t \dot{X}_{it}\dot{Y}_{it}) \\ &= (\sum_i \dot{X}_i'\dot{X}_i)^{-1}(\sum_i \dot{X}_i'\dot{Y}_i)\end{aligned}$$

显然 β_{fe} 是无偏的，而 $V_{fe}^0 = \sigma_\epsilon^2(\sum_i \dot{X}_i'\dot{X}_i)^{-1}$

3.4 差分估计

除了采用均值处理，同样也可用差分处理，易得：

$$\Delta Y_{it} = \Delta X_{it}'\beta + \Delta\epsilon_{it}$$

而 $\Delta Y_i = D_iY_i$。如果 $T = 2$，那么 $\hat{\beta}_\Delta = \hat{\beta}_{fe}$，但是所得并非同方差，由于 D_iD_i' 的结果，因此也需要 GLS（同除 $H_i^{-\frac{1}{2}}$）。

3.5 Dummy Variables Regression

目前不作要求

Chapter 4

Lecture 8 IV

4.1 IV 概述

我们对于内生性 (endogenous) 的定义为 $E[Xe] \neq 0$，等价的 $E[Xe] = 0$ 是外生性。一个内生性回归总是可以拆解为含内生性和不含内生性的部分，也即

$$Y = X_1'\beta_1 + X_2'\beta_2 + e$$

其中 $X = (X_1, X_2)$ 而 X_1 外生 X_2 内生。在本节我们将 X_2 命名为 Y_2，而 Y 命名为 Y_1。

那么什么是工具变量呢？我们称，一个 $l \times 1$ 的列向量 Z 是工具变量，当这个变量满足以下几个条件：

$$E[Ze] = 0 \quad E[ZZ'] > 0 \quad \text{rank}(E[ZX']) = k$$

第一个保证了外生性，第二个排除了线性冗余性，第三个是相关性条件（relevance condition），也是该期望可逆的条件。由于原变量 X_1 天然具有外生性，因此不妨令 $Z = (X_1, Z_2)^T$（不然哪里找那么多工具变量！）。显然当 Z 的维度等于 k 时这个模型是 just identitied（唯一性），超过就不唯一了（over identitied）。

为了更好地描述内生变量与工具变量的关系，我们不妨用一个回归来表述它：

$$Y_2 = \Gamma' Z + u_2 = \Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2$$

显然,$E[Zu_2] = 0$。我们把上述关系代入到内生性回归中可以得到如下结果：

$$\begin{aligned} Y =& Z_1'\beta_1 + (\Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2)'\beta_2 + e \\ =& Z_1'(\beta_1 + \Gamma_{12}\beta_2) + Z_2'\Gamma_{22}\beta_2 + u_2'\beta_2 + e \end{aligned}$$

记 $\lambda = \bar{\Gamma}\beta$，根据上式的结果我们可以得 $\bar{\Gamma}$ 为

$$\begin{bmatrix} I_{k1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix}$$

（太显然了）通过 Y_1, Y_2 的关系我们可以计算出最小二乘估计。下面我们来确定各自的维度。已知都是列向量，$X_2(Y_2)$ 是 k_2 维，Z 是 l 维，那么显然 Γ 是 $l \times k_2$ 维。因此，整个 $\bar{\Gamma}$ 是 $l \times k$ 维的，因为 $Y(Y_1)$ 是 k 维的。同时由于 $l = k_1 + l_2$，那么也可确定各自的维度。下面我们来研究为什么 rank$(E[ZX']) = k$。首先我们可以很由 Γ 和 λ 的投影系数得到下述关系：

$$\Gamma = E[ZZ']^{-1}E[ZX_2'] \tag{4.1}$$

$$\lambda = E[ZZ']^{-1}E[ZY_1'] \tag{4.2}$$

$$= E[ZZ']^{-1}E[ZX']\beta \tag{4.3}$$

同样的，$\bar{\Gamma} = E[ZZ']^{-1}E[ZX']$（因为残差为 0）故 β 要是有解，矩阵 $E[ZX']$ 必须可逆，且当 β 由唯一解时，矩阵 $E[ZX']$ 的秩等于 $E[ZY_1']$，而 $E[ZY_1']$ 应是 k 维的（因为 Y_1 是 k 维）。如果 $l = k$，会发生什么？这时有唯一解，因此 $E[ZX']\beta - E[ZY_1'] = 0$（这也可从 $E[Ze] = 0$ 推出，因为 $Y_1 - X'\beta = 0$。则 $\hat{\beta}_{iv}$ 可计算出（具体值是多少？）。实际上，由于 $\lambda = \bar{\Gamma}\beta$ 也可计算出同样的结论（把上方的式子代入），这时又称之为 indirect least Squares (ILS) estimator。所以一共有两种方法推导 β。这都需要掌握。对于有 $Y_1 = X'\beta + \alpha + e$ 这类回归，我们牢记 $\hat{\alpha}_{IV} = \bar{Y}_1 - \bar{X}'\hat{\beta}_{IV}$ 可得。（这部分可以回顾一下，我觉得很丑陋，所以仅仅简单了解了一下）。

4.2 WALD estimator

（要加强记忆）在 Wald 估计量中，我们是单变量，IV 是二元变量。Wald 估计量有两个重要的式子：

$$E[Y|Z=1] = E[X|Z=1]\beta + \alpha \tag{4.4}$$

$$E[Y|Z=0] = E[X|Z=0]\beta + \alpha \tag{4.5}$$

$$\tag{4.6}$$

由此可得：

$$\beta = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}$$

自然的，由 $\bar{Y}_1 = \frac{\sum_i Z_i Y_i}{\sum_i Z_i}$ 可以得到 Wald 估计量 $\hat{\beta} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$。那么，如何证明 IV 估计量 $\hat{\beta}_{IV} = \frac{\bar{Y}_1 - \bar{Y}}{\bar{X}_1 - \bar{X}}$ 就是 Wald 估计量呢？我们从 $\bar{Y} = Z_1\bar{Y}_1 + (1 - Z_1)\bar{Y}_0$ 得出。

4.3 Two-Stage Least Squares

令 $W=\bar{\Gamma}'Z$，可得到关于 W 的估计，然后把 W 代入可得到 β 关于 $\bar{\Gamma}$ 的。如果我们同时用之前所得的 $\bar{\Gamma}=E[ZZ']^{-1}E[ZX']$，那么将其代入可得到 $\hat{\beta}_{2sls}$（其实以 $(Z'Z)^{-1}(Z'X)$ 代入）:

$$\hat{\beta}_{2sls}=(X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y_1$$

注意到，这和投影矩阵类似，令 $P_Z=Z(Z'Z)^{-1}Z'$，以及已知的投影矩阵的性质，可得到如下结果

$$\hat{\beta}_{2sls}=(X'P_ZP_ZX)^{-1}X'P_ZY_1=(\hat{X}'\hat{X})^{-1}\hat{X}'Y_1$$

注意到，两阶段最小二乘同样满足大数定理。要证明其收敛到 β，只需将其中 Y_1 写成 $X\beta+e$ 形式

4.4 局部平均处理效应（Local Average Treatment Effects）

令 $Y(x)=h(x,U),X(z)=g(z,U)$，以观察 $X=X(z),Y=Y(X)$。因此平均处理效应可定义为

$$\text{ATE}=E[C]=E[Y(1)-Y(0)]$$

但是这并不能很好地在内生性情形下实现，一个具体的列子是上学。由于假设 $X(0)=1,X(1)=0$ 的 “Defiers” 不存在，而 never takers 和 always takes 之 $X(1)-X(0)=0$ 恒成立，因此研究 $E[X(1)-X(0)]$ 往往只针对”compliers” 生效，也即

$$\text{LATE}=E[Y(1)-Y(0)|X(1)>X(0)]$$

等价于 Wald 估计量。推导如下:

Proof. 首先记 $Y(1)-Y(0)=C$，按照概率分布的形式，有:

$$X=(1-Z)X(0)+ZX(1)=X(0)+Z(X(1)-X(0))$$

将其代入 Y 中:

$$Y=Y(0)+X(Y(1)-Y(0))=Y(0)+XC=Y(0)+X(0)C+Z(X(1)-X(0))C$$

那么期望可表述为:

$$E[Y|Z=1]=E[Y(0)]+E[X(0)C]+E[(X(1)-X(0))C]$$
$$E[Y|Z=0]=E[Y(0)]+E[X(0)C]$$

则

$$
\begin{aligned}
E[Y|Z=1]-E[Y|Z=0] &= E[(X(1)-X(0))C] \\
&= 1\times E[C|X(1)-X(0)=1]P[X(1)-X(0)=1]+0\times E[C|X(1)-X(0)=0] \\
&= E[C|X(1)-X(0)=1](E[X|Z=1]-E[X|Z=0])
\end{aligned}
$$

可从 $P[X(1)-X(0)=1]=E[X(1)-X(0)]$ 得。那么，证毕。 $\square$

Chapter 5

Lecture 9 DID

5.1 名词解释

1. 一次性政策：干预一次
2. 连续性政策：一段时间内持续进行的政策干预或政策措施
3. 平行趋势假设：在没有干预的情况下，处理组（treated group）和对照组（control group）的趋势是相同的
4. 处理效应：干预或政策的实施，对处理组产生的影响，是针对某个个体的，因此期数没有显示写出：
$$C_i = Y_i(1) - Y_i(0)$$
5. 预期效应：个体或群体由于预期到未来某个政策或干预的实施，而在实际干预前已经开始作出反应。预期效应是内生性问题的一种表现
$$Y_{i1}(1) = Y_{i1}(0)$$
6. 平行趋势：平行趋势是指在干预发生之前，处理组和对照组的因变量（如收入、消费等）在时间上的变化趋势是一致的。也即：
$$E[Y_{i2}(0) - Y_{i1}(0)|D_i = 1] = E[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]$$
7. ATT：指干预对那些实际接受干预的个体或单位的平均处理效应。
$$\begin{aligned}&E[Y_{i2}(1) - Y_{i2}(0)|D_i = 1]\\&= (E[Y_{i2}|D_i = 1] - E[Y_{i1}|D_i = 1]) - (E[Y_{i2}|D_i = 0] - E[Y_{i1}|D_i = 0])\end{aligned}$$

这也是 DID 的真实含义若再基于 PO 假设：

$$E[Y_{i2}(1) - Y_{i1}(1)|D_i = 1] - E[Y_{i2}(0) - Y_{i2}(0)|D_i = 0]$$

若再基于无预期假设

$$E[Y_{i2}(1) - Y_{i1}(0)|D_i = 1] - E[Y_{i2}(0) - Y_{i2}(0)|D_i = 0]$$

加减 $E[Y_{i2}(0)|D_i = 1]$ 可以是第一项变成 ATT，而基于平行趋势假设后面的可以消去

8. ATF：指干预对所有人群的平均处理效应。

9. TWFE：双向固定效应模型，广泛用于面板数据分析中。该模型同时控制了个体固定效应（例如，公司、地区或个人的时间不变特征）和时间固定效应（例如，所有个体在同一时间点受到的共同冲击）。

10. Potential Outcomes：假设每个个体在不同的处理状态下都有一个潜在的结果。然而，由于每个个体只能处于一个状态（接受或不接受处理），因此只能观察到一个潜在结果。

11. 检验平行趋势：如果事前趋势一致，那么假设平行趋势一致（但无法被检验），一般可视化。也可用一回归检验（怎么写呢）

5.2 DID 前沿理论

Staggered DID

一次性政策发生在多个不同的时间，这也成为异质性处理效应，无法被双向固定效应解决（因为存在内生性）。此时的平行趋势为：

$$E[Y_{i,t}(\infty) - Y_{i,t-1}(\infty)|G_i = g] = E[Y_{i,t}(\infty) - Y_{i,t-1}(\infty)|G_i = g']$$

∞ 表示未被处理。此时的无预期效应为：

$$Y_{it}(\infty) = Y_{it}(g) \quad t < g$$

相当于被处理前都是一致的。

5.2.1 Forbidden Comparsion

当处理效应并非常数（也即 $Y_{it}(g) - Y_{it}(\infty)$ 不定，随时间变化），可能会导致负权重问题，是因为采用了“forbidden comparsion”。有一个简单的例子描绘此类情形，如果存在一个组，两个期都是被处理的（always-treated），一个组只在第二期被处理 (swithers)，那么估计系数，或是平均处理效应为：

$$\hat{\beta} = (\bar{Y}_{s,2} - \bar{Y}_{s,1}) - (\bar{Y}_{at,2} - \bar{Y}_{at,1})$$

负权重产生的原因在于 AT 组随时间增大！

5.2.2 Frisch-Waugh-Lovell theorem

是一个并非局限于 DID 的定理，在多元回归分析中，先对一些自变量进行回归，再进行剩余部分的回归，得到的回归系数与直接对全部自变量进行回归时的结果是相同的。那么，在 DID 中是如何体现的呢？首先，我们有这样的回归方程：

$$Y_{it} = \alpha_i + \phi_t + D_{it}\beta + \epsilon_{it}$$

我们将 D_{it} 对其余变量进行回归，在将 D_{it} 对回归后的残差 $D_{it} - \hat{D}_{it}$ 进行回归，得到的估计系数 $\beta = \frac{E[Y_{it}(D_{it}-\hat{D}_{it})]}{Var(D_{it}-\hat{D}_{it})}$ 即为直接回归的估计系数。但是是什么导致了负系数的出现？可以从分布回归中揭示。显然，β 只在第二步回归中出现，因此 β 的正负必然与残差项有关

5.2.3 SA

上述几节探讨的前沿理论都是围绕 DID，或是 TWFE 中的负权重问题。在本节中，一个新的问题被提出，也即即使不是负的，依然可能是不恰当的。要仔细说明这个情况，先来看公式：

$$ATT(g,t) = E[Y_{it}(g) - Y_{it}(\infty)|G_i = g]$$

注意到 g 是处理时点，因此 $t = g - 1$ 同样满足形式。请推导此情形下逐步采用 PO 和无预期后如何等价于 ATT 的过程。虽然计算过程显示是合理的，但是在计算均值时，如果样本不平衡，也可导致偏误。这些写的很不清楚，但就 GPT 而言，SA 更多是用于处理滞后效应，也即对 $ATT(g, g+k)$，如果 $k < 0$ 那么显然是存在事前趋势的。

Chapter 6

Lecture 9 Regression Discontinuity

6.1 The Sharp Regression Discontinuity Design

1. $W_i = 1\{X_i \geq c\}$

2.

$$\begin{aligned} E[Y|X=x] = & E[Y|W=0, X=x]P(W=0|X=x) \\ & + E[Y|W=1, X=x]P(W=1|X=x) \end{aligned}$$

3. average causal effect of the treatment at the discontinuity point:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c]$$

4. conditional unconfoundedness assumption, overlap assumption

5. overlap assumption 被假设 conitnuity of conditional regression functions assumption 所补偿，此时的 τ_{SRD} 为?

6.

$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \mid X = x]}{\lim_{x \downarrow c} \mathbb{E}[W \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[W \mid X = x]}$$

7. 为什么 Fuzzy RD is IV? 写出 complier, never-takers 和 always-takers 的形式

6.2 Graphical Analysis

1. 绘制 Y 在断点附近的区间的直方图

2. 比较其他协变量在断点附近的均值，比如 W_i

3. 线性回归，对远离点敏感

4. 核回归，核相减

5. 局部线性回归，按照距离 c 点的距离

Chapter 7

不敢忘，不能忘（发出兔友的声音）

7.1 Lecture 1

1. 有哪五种数据类型（6）
2. 什么是 CEF（13）
3. 迭代期望法则有哪些（15）
4. 什么是 CEF error，具有哪些性质？（16）
5. unconditional variance **的大小解释了什么？(17)**
6. 什么是 best predictor？谁是 best predictor？为什么？（18）
7. 什么是 conditional error variance?（20）
8. 什么是交互影响？（23）
9. 什么是 best linear predictor，它的假设是什么？（24）
10. 什么是 linear projection coefficient？（25）
11. projection error 具有怎样的性质？（27）
12. 推导 linear predictor error variance（28）
13. 推导 regression coefficients with intercept（29）
14. 推导 omitted variable bias（30），并说明什么是 omitted variable bias
15. linear CEF 和 best linear predictor 的区别是什么

7.2 Lecture 2

1. OLS estimator，以及什么对应样本矩，什么对应总体矩
2. 什么是残差，如何写出？误差是什么。**残差满足什么样的性质**
3. demand regressors 是什么？这个回归中 OLS estimator 等于多少
4. 矩阵形式下 OLS estimator 是多少
5. 投影矩阵的形式是什么？写出它与 X,Y，均值以及自身的组合形式
6. 幂等矩阵的形式是什么？写出它与 X,Y，均值以及自身的组合形式
7. **误差的方差的自然估计，可行估计是什么**？它们与幂等矩阵有什么关系，谁更小？
8. SST，SSE，SSR **的形式，**R^2 **的形式**
9. 两阶段最小二乘，如何**通过残差进行简易计算**？
10. Frisch-Waugh-Lovell **定理，在多变量下与上面的结果有什么联系**？
11. h_{ii} **等于什么？这种回归与 OLS 估计存在什么联系**
12. 如何判断一个观测值的影响力

7.3 Lecture 3

1. 线性回归的条件（满秩）
2. 证明 OSL estimator 是无偏的，以及迭代期望法则下的
3. 什么是条件协方差矩阵？D 是什么？什么情况下下 D 和 σ 有简洁的联系？
4. $V_{\hat{\beta}}$ 是什么？化成含 σ 形式
5. **解释何为高斯-马尔可夫定理**（5）？什么是 BLUE？同方差下证明 OLS estimator 是 BLUE
6. 广义最小二乘下有什么不一样的条件？这时的 $\hat{\beta}$ 是无偏的吗？计算此时的 $V_{\hat{\beta}}$
7. 写出 $\tilde{\beta}_{gls}$

8. **残差的条件协方差是什么？**
9. **误差方程的估计是什么？其条件期望是什么？为什么要用 s^2，其具有怎样的性质？**
10. **什么是 $V_{\hat{\beta}}^0, \hat{V}_{\hat{\beta}}^0$，后者的条件期望是什么？**
11. 什么是 ideal but infeasible estimator?
12. 什么是 $\hat{V}_{\hat{\beta}}^{HC0}$? HC0 指什么?
13. 什么是 $\hat{V}_{\hat{\beta}}^{HC1}$? HC1 指什么?
14. 解释 $\rho^2, R^2, \bar{R}^2$ 的联系

7.4 Lecture 4

1. 正态回归模型具有怎样的特征？当什么情况下成立？（9）
2. 写出似然函数，并求出对应系数（12）
3. 写出估计系数与残差的分布形式（15）
4. 写出 variance estimator 的分布形式（17）
5. 写出 t-statistic 的形式（18），并写出 t-ratio 形式（19），谈论这些统计量适宜的条件（20）

7.5 Lecture 5

1. **两种收敛，三种定律，一个方法**
2. $\hat{\beta}$ 的收敛是怎样的
3. 什么是渐进正态？Ω 是什么？谁和它有关系？
4. 什么是 $V_\beta, V_{\hat{\beta}}, V_\beta^0$，它们的关系是怎么样的？
5. $\hat{\sigma}^2, s^2$ 如何收敛？
6. 什么是 $\hat{V}_\beta^0, \hat{V}_\beta^{HC0}, \hat{\Omega}, \hat{V}_{\hat{\beta}}^{HC0}$，它们的关系是怎样的？仅以 $\beta, 0$ 为例，一共有四种组合，分别是怎样的

7. 什么是 asymptotic distribution of functions of parameters（29）
8. 写出此时的 $T(\theta)$ 以及收敛情形（33）
9. 写出如果估计 $m(x)$(38)

7.6 Lecture 6

1. 说明 regression model, projection model, structural model 的区别
2. 什么情况下内生性不会发生？内生性偏误由什么导致？
3. 举出三种引起内生性偏误的例子

7.7 Lecture 7

1. 简单描述这几个概念：微观面板，宏观面板，平衡与非平衡面板
2. 写出混合回归的估计系数，分别矩阵形式和标量形式，并给出其一致性的条件
3. 分别写出 one-way error componet model 的含 t 和不含 t 形式
4. 写出随机效应模型的六个条件
5. 写出此时误差项的条件方差-协方差矩阵，此时的估计系数。相较于混合回归，谁的方差更小？
6. 什么是固定效应？什么情况下会导致混合回归和随机效应模型出现偏差，**这是由什么造成的**？
7. 固定效应模型的假设是什么？
8. **写出固定效应模型转化后的形式**，谈论它的后果，并写出**它的估计系数、方差**
9. 差分估计系数中 D_i 长什么样？
10. 写出此时的两种回归方程，以及估计系数。什么情况下等价于固定效应模型？
11. 协方差矩阵长什么样？写出**广义线性回归下的一系列系数**
12. 简单解释哑变量回归的含义