Econometrics(II) Cheatsheet

§ Time Series Data

Definition: data collected on the same observational unit at multiple time periods

- dynamic causal effect: $\Delta X_t \to Y_{t+1}, Y_{t+2}, \cdots Y_{t+k}, \cdots$
- economic forecasting: best forecast of $Y_{T+1}, Y_{T+2}, \cdots$
- modeling risks in financial market: variances pattern, volatility clustering

Time Series Data: a realization of a stochastic/random process (indexed by time)

Notations: 1. jth lag: Y_{t-j} (Stata: L.y, L2.y, ...) 2. jth lead: Y_{t+j} (Stata: F.y, F2.y, ...)

3. jth difference: $\Delta Y_{t-j} = Y_t - Y_{t-j}$ (Stata: D.y) $\Rightarrow \Delta \ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1})$ (percentage change: $\approx 100\Delta \ln(Y_t)$%)

4. autocovariance: $Cov(Y_t, Y_{t-j})$; autocorrelation: $Corr(Y_t, Y_{t-j})$ $\Big\}$ $Corr(Y_t,Y_{t-j}) = \frac{Cov(Y_t,Y_{t-j})}{\sqrt{Var(Y_t)Var(Y_{t-j})}} = \rho_j$ (jth autocorrelation (serial correlation) coefficient), describe population joint distribution (j)

jth sample autocorrelation $\hat\rho_j = \frac{\widehat{Cov}(Y_t, Y_{t-j})}{\widehat{Var}(Y_t)}$, $\widehat{Cov}(Y_t, Y_{t-j}) = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \bar Y_{j+1,T})(Y_{t-j} - \bar Y_{1,T-j})$, $\hat\rho_j = \rho(j)$

ACF (Auto Correlation Function) plot / correlogram (Stata: corrgram)

1 Basic Models and CLM (classical linear model) for time series

- Static Model: $Y_t = \beta_0 + \beta_1 Z_t + U_t$ (contemporaneous causal effect)
- FDL (Finite Distributed Lag) Model: $Y_t = \alpha_0 + \delta_0 Z_t + \delta_1 Z_{t-1} + \delta_2 Z_{t-2} + \cdots \delta_r Z_{t-r} + U_t$
  - δ_0: impact propensity, δ_r: r-period dynamic multiplier
  - $\delta_0 + \delta_1 + \delta_2$: 2-period cumulative dynamic multiplier
  - a temporary change (ΔZ_t): $\Delta Y = Y_t - Y_0$: $0, \delta_0, \delta_1, \delta_2, \cdots \delta_r, 0$ (impact propensity / dynamic multipliers)
  - a permanent change (ΔZ_{t+1}): $\Delta Y = Y_t - Y_0$: $0, \delta_0, \delta_0+\delta_1, \delta_0+\delta_1+\delta_2, \cdots (\delta_0 + \cdots \delta_r), \cdots$
  - $\Rightarrow$ cumulative / long-run propensity (LRP) — LRP $= \delta_0 + \delta_1 + \cdots \delta_r$ here

Assumption TS.1 (Linear in Parameters): $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots \beta_k X_{tk} + U_t$

Assumption TS.2 (No Perfect Collinearity): $X_1 \neq$ linear combination $(X_2, X_3, \cdots X_K)$ (两边同时减去mean)

Assumption TS.3 (Zero Conditional Mean): $E(U_t|X) = 0$, $X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & & & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{Tk}\end{bmatrix} \Rightarrow \vec X_2$

"mean independence implies" 1. $E(U)=0$, 2. $Cov(X_t, U) = 0$, or: $E(U_t X) = 0$

- $E(U_t|X)$: strictly exogenous, $t = 1, 2, \ldots, T$
- $E(U_t|\vec X_t)$: contemporaneously exogenous

(The essence of the SLR.2 Random Sampling in CLM is to ensure this) stable volatility over time

- unbiasedness of OLS: $E(\hat\beta_j) = \beta_j$, $j = 0, 1, 2, \ldots, k$

Assumption TS.4 (Homoskedasticity): $Var(U_t|X) = Var(U_t) = \sigma^2$, $t = 1, 2, \ldots, T$

Assumption TS.5 (No Serial Correlation): $Corr(U_t, U_s|X) = 0$, $t \neq s$ 可以替代SLR.2

(Random Sampling $+ E(U_t|\vec X_t) \Rightarrow E(U_t|X) = 0$, also $\Rightarrow Corr(U_t, U_s|X) = 0$)

- OLS sampling variance: $Var(\hat\beta_j|X) = \sigma^2 / SST_j(1 - R_j^2) = \sigma_u^2 / [T\sigma_j^2 (1-R_j^2)]$
  unbiased estimation of error variance: $E(\hat\sigma^2) = \sigma^2 = E(\hat u_i^2/df) = E(SSR/df)$
- Gauss-Markov Theorem: OLS estimators are BLUE

Assumption TS.6 (Normality): $U_t \sim Normal(0, \sigma^2)$, independent of X

- OLS estimator ~ Normal Distribution conditional on X, F test, T test... ✓

§ Modelling a linear time trend: $Y_t = \alpha_0 + \alpha_1 t + e_t \Rightarrow \frac{dY}{dt} = \alpha_1$

Modelling an exponential time trend: $\log(Y_t) = \alpha_0 + \alpha_1 t + e_t \Rightarrow \frac{d\log(Y)}{dt} = \alpha_1$

spurious relationship: reg trending variables on each other solely $\Rightarrow Cov(X, \text{trending factors in } U) \neq 0$

Detrended Series: $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots \beta_k X_{tk} + \alpha t + U_t$

$\Rightarrow$ detrend Y_t, X_{tj} and get $\ddot y_t = Y_t - \hat\alpha_0 - \hat\alpha_1 t$, $\ddot x_{tj}$, and reg $\ddot y_t$ on $\ddot x_{tj}$

[R^2 analysis] $R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\hat\sigma_u^2}{\hat\sigma_Y^2}$, $\bar R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat\sigma_u^2}{\hat\sigma_Y^2}$

(even a time trend is included in the regression, SST is over estimated, thus R^2 is boosted.)

spurious regression: $SST(\hat Y)\uparrow \Rightarrow R^2, \bar R^2$ overestimate!

detrended regression: $R^2 = 1 - SSR/SST'$ ($SST' = \sum \ddot y_t^2$), $df(\hat\sigma) = n-2 \Rightarrow \hat\sigma_Y^2 = \frac{SST'}{n-2}$ (trend parameters)

(other): Seasonality and deseasonalizing (add season dummies)

2 Further Issues with Time Series Data and OLS

- § strictly stationary: $\forall m, \forall h$, joint distribution $(X_{t1}, \ldots X_{tm}) = (X_{t1+h}, \ldots X_{tm+h})$
  | weak / covariance stationary: 1. $E(X_t) = \mu$ 2. $Var(X_t) = \sigma^2 \ \forall t$, $Cov(X_t, X_{t+h}) = f(h)$
  - $\Rightarrow$ ACF (Auto Correlation Function): $\rho_h = Corr(Y_t, Y_{t+h})$
  - PACF (Partial Auto Correlation Function): β_h in $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_h Y_{t-h} + e_t$
- weakly independent: Covariance stationary $+ \rho(h) \to 0$ as $h \to +\infty$
  (asymptotically uncorrelated) $\Rightarrow$ LLN and CLT hold true

(1) White noise process: $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_{t-j}) = 0$, $Var(\varepsilon_t) = \sigma^2 \Rightarrow$ ACF=0, weakly dependent
  - § independent W.N.: ε_t's are independent
  - Gaussian W.N.: $\varepsilon_t \sim N(0, \sigma^2)$

(2) MA(1): Moving Average process of order one: $Y_t = C + \varepsilon_t + \theta\varepsilon_{t-1}$ (ε_t i.i.d.)

(3) AR(1): AutoRegressive process of order one: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$, $|\beta_1| < 1$
  - MA(1): ACF $= \frac{\theta}{1+\theta^2}$ ($h=1$), 0 ($h \geq 2$); AR(1): ACF $= \beta_1^h \Rightarrow$ weakly dependent

(4) RW / Random Walk: AR(1)($\beta_1 = 1$): highly persistent, strongly dependent, non-stationary

Assumption TS.1* (Linear in Parameters): TS.1 + Stationary + Weakly dependent (both Y and X)

Assumption TS.2* (No perfect Collinearity): TS.2

Assumption TS.3* (Zero conditional mean): $E(U_t|\vec X_t) = 0$ (contemporary)

- consistency for $\hat\beta_j$: $\hat\beta_j \xrightarrow{p} \beta_j$ ($E(\vec X_t U_t) = 0$ would be suffice)

Assumption TS.4* (Homoskedasticity): $Var(U_t|\vec X_t) = Var(U_t) = \sigma^2$ (contemporary)

Assumption TS.5* (No Serial Correlation): $Corr(U_t, U_s|\vec X_t, \vec X_s) = 0$

- OLS estimator: asymptotically normally distributed

(※) time series with deterministic time trend: nonstationary
  trend-stationary: weakly dependent + stationary when detrended

Random Walk: $Y_t = Y_{t-1} + \varepsilon_t = \cdots = \varepsilon_t + \varepsilon_{t-1} + \cdots \varepsilon_1 + Y_0$ (ε_t: i.i.d.)

- § nonstationary: $E(Y_t) = E(Y_0)$, $Var(Y_t) = t\sigma_\varepsilon^2 \Rightarrow$ trending data
  | no weakly independent: $Corr(Y_t, Y_{t+h}) = \sqrt{t/(t+h)}$ (depends on t)
- (※) special case of Unit Root Process: ε_t is weakly independent

1. Random Walk with drift: $Y_t = \alpha_0 + Y_{t-1} + \varepsilon_t = \cdots = \alpha_0 t + \varepsilon_t + (\varepsilon_{t-1} - \varepsilon_1) + Y_0$
   $E(Y_t) = \alpha_0 t + E(Y_0)$, $Var(Y_t) = t\sigma_\varepsilon^2$, $Corr(Y_t, Y_{t+h}) = \sqrt{t/(t+h)}$ (R.W.d has both "trends")

2. § drift: deterministic trend
   | stochastic trend: propensity to trend due to increasing variance

3. order of integration
   - I(0): weakly independent time series
   - I(1): has to be differenced once to get weakly dependent series
     (e.g.) $Y_t = Y_{t-1} + \varepsilon_t \Rightarrow \Delta Y_t = \varepsilon_t \Rightarrow$ weakly dependent
     Test for I(1): unit root test $\Leftarrow \hat\rho_1 = \widehat{Corr}(Y_t, Y_{t-1}) \sim 1$
     or it may have a deterministic trend!!

4. Dynamically Complete Model: enough lagged variables included
   $E(Y_t|X_t, Y_{t-1}, X_{t-1}, Y_{t-2}, \cdots) = E(Y_t|X_t) \Rightarrow$ no serial correlation (not necessary; correct causal effect (though) serial correlation)
   - Sequential Exogeneity: enough lagged X variables included
     $E(U_t|X_t, X_{t-1}, \cdots) = E(U_t) = 0$ (weaker than $E(U_t|X) = 0$)
   - (※) Sequential Exogeneity + lagged Y variable (as explanatory variable) = DCM

3 Test for Serial Correlation

- § OLS not BLUE, statistic testing invalid
  | R^2 still work given that data are stationary and weakly dependent

AR(1) with strictly exogenous regressors: $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots \beta_k X_{tk} + U_t$, $E(U_t|X) = 0$ (AR(1) model for serial correlation)

$\Rightarrow U_t = \rho U_{t-1} + \varepsilon_t$ (ε_t i.i.d.), test $\hat u_t = \rho \hat u_{t-1}$ + error

H_0: $\rho = 0$ (no serial correlation) $\Rightarrow$ t-test valid asymptotically

- Durbin-Watson Test (TS.1-TS.5)

$$DW = \frac{\sum_{t=2}^{T}(\hat u_t - \hat u_{t-1})^2}{\sum_{t=1}^{T}\hat u_t^2} \approx 2(1-\hat\rho)$$

H_0: $\rho = 0$ (no serial correlation), H_1: $\rho > 0$. Reject H_0 if $DW < d_L$, Fail to reject if $DW > d_U$ (Stata: estat dwatson)

- if X not strictly exogenous:

AR(1): $\hat u_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \cdots + \alpha_k X_{tk} + \rho\hat u_{t-1} + \text{error}$ (H_0: $\rho = 0$)

AR(q): $\hat u_t = \alpha_0 + \alpha_1 X_{t1} + \cdots \alpha_k X_{tk} + \rho_1 \hat u_{t-1} + \rho_2 \hat u_{t-2} + \cdots \rho_q \hat u_{t-q} + \text{error}$ (H_0: $\rho_1 = \rho_2 = \cdots \rho_q = 0$)

- F, Wald, LM test ∘ Breusch-Godfrey test: $(n-q)R^2$, the LM version ($\hat u_t$, $E(U_t|X) = 0$)

- Q Test for Serial Correlation ($p = 1, 2, 3, \cdots \Leftarrow$ test for simultaneous $\rho = 0$) $\Rightarrow$ asy. distribution in reality

sample jth order autocorrelation $\hat\rho_j = \hat r_j / \hat r_0$, where $\hat r_j = \frac{1}{n}\sum_{t=j+1}^{T}(U_t - \bar U_n)(U_{t-j} - \bar U_n)$

- ∘ Box-Pierce Q test: $Q_n \equiv n\sum_{j=1}^{p}\hat\rho_j^2 = \sum_{j=1}^{p}(\sqrt n \hat\rho_j)^2 \xrightarrow{d} \chi^2(p)$, suppose $\{U_t\}$ i.i.d $(0, \sigma^2)$
  (※) $\sqrt n \hat r \xrightarrow{d} N(0, \sigma^4 I_p)$, $\sqrt n \hat\rho \xrightarrow{d} N(0, I_p)$, where $\hat r$ and $\hat\rho$ are vectors of $\hat r_j$ and $\hat\rho_j$
- ∘ Ljung-Box's modified Q test: $Q \equiv n(n+2)\sum_{j=1}^{p}\hat\rho_j^2/(n-j) = \sum_{j=1}^{p}\frac{n+2}{n-j}(\sqrt n \hat\rho_j)^2 \xrightarrow{d} \chi^2(p)$
  better approximation for moderate sample size — $Q_n^{(2)} \equiv n\hat\rho'(I_p - \hat\Phi)^{-1}\hat\rho \xrightarrow{d} \chi^2(p)$
- ∘ if $\{X_t\}$ not strictly exogenous
  $\begin{cases} E(U_t|U_{t-1}, U_{t-2}, \cdots, X_t, X_{t-1}, \cdots) = 0 \\ E(U_t^2|U_{t-1}, U_{t-2}, \cdots, X_t, X_{t-1}, \cdots) = \sigma^2 \end{cases} \Rightarrow \sqrt n \hat r \xrightarrow{d} N(0, \sigma^4(I_p - \Phi))$, $\sqrt n \hat\rho \xrightarrow{d} N(0, I_p - \Phi)$
  with $\phi_{jl} = E(\vec X_t U_{t-j})' E(\vec X_t \vec X_t')^{-1} E(\vec X_t U_{t-l}) / \sigma^2$

4 Correcting for Serial Correlation

assumptions: all Gauss-Markov Assumption except TS.5

Assume errors (U_t) follows AR(1): $U_t = \rho U_{t-1} + \varepsilon_t$, $Var(U_t) = \frac{\sigma_e^2}{1-\rho^2}$, e_t: i.i.d.

Quasi-differencing (FGLS approach): $\begin{cases} Y_t = \beta_0 + \beta_1 X_t + U_t \\ \rho Y_{t-1} = \beta_0\rho + \rho\beta_1 X_{t-1} + \rho U_{t-1} \end{cases} \Rightarrow (Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + e_t$ (required TS.3 for consistency)

1. reg Y_t on $1, X_{t1}, \cdots X_{tk} \Rightarrow$ obtain $\hat u_t$ (Omit the first observation)
2. reg $\hat u_t$ on $1, \hat u_{t-1} \Rightarrow$ obtain $\hat\rho$ (AR(1) assumption)
3. reg $\tilde y_t (= Y_t - \hat\rho Y_{t-1})$ on $1, (X_t - \hat\rho X_{t-1}) \Rightarrow$ efficient($\hat\beta$), not unbiased se: $\Rightarrow$ consistent but biased

CO (Cochrane-Orcutt); PW (Prais-Winsten) adds a transformed first observations

- HAC (Heteroskedasticity and AutoCorrelation Consistent)
  OLS: $Y_t = \beta_0 + \beta_1 X_{t1} + \cdots \beta_k X_{tk} + U_t$, SER $\hat\sigma = \sqrt{\frac{SSR}{n-k-1}}$, $SSR = \sum_{t=1}^{T}\hat u_t^2$
  auxiliary: $X_{t1} = \delta_0 + \delta_2 X_{t2} + \cdots \delta_k X_{tk} + r_t$, AVar($\hat\beta_1$) be estimated from $\hat u_t$, $\hat r_t$
  $\Rightarrow AVar(\hat\beta_1) = [\sum_{t=1}^{T}E(r_t^2)]^{-2} Var(\sum_{t=1}^{T} r_t U_t)$ Newey-West formula: $\hat a_t = \hat r_t \hat u_t$
  $\Rightarrow se(\hat\beta_1) = [\text{"}se(\hat\beta_1)\text{"}/\hat\sigma]^2/\sqrt{\hat v}$, where $\hat v = \sum_{t=1}^{T}\hat a_t^2 + 2\sum_{h=1}^{g}[1 - \frac{h}{g+1}](\sum_{t=h+1}^{T}\hat a_t \hat a_{t-h})$
  (the usual OLS se are normalized and inflated by serial correlation)
  (※) g controls how much serial correlation is allowed ($g = 1, 2, \cdots$)
- First test for Serial Correlation: $\hat u_t = \alpha_0 + \beta_1 X_t + \delta\hat u_{t-1} + \text{error}$, after correction
  Then test for heteroskedasticity: $\hat u_t = \alpha_0 + \beta_1 X_t + \text{error}$

5 ARIMA models

AutoRegressive Integrated Moving Average (p, d, q)

- autoregression of order p: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_p Y_{t-p} + \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$
  - AR(1): $Y_t = \beta_0/(1 - \beta_1 L) + \varepsilon_t/(1 - \beta_1 L) \Rightarrow E(Y_t) = \beta_0/(1-\beta_1)$, Y_t stationary if $|\beta_1| < 1$
  - AR(2): $(1 - \beta_1 L - \beta_2 L^2)Y_t = \beta_0 + \varepsilon_t = (1 - \lambda_1 L)(1 - \lambda_2 L)Y_t$, Y_t stationary if $|\lambda_1| < 1$ and $|\lambda_2| < 1$
    - one root $= 1 \Rightarrow$ I(1): first difference $\to$ stationary
    - two roots $= 1 \Rightarrow$ I(2): second difference $\to$ stationary
    - ($\beta_1 + \cdots \beta_p < 1 \Rightarrow$ stationary)
  - AR(p): $\mu = E(Y_t) = \beta_0/(1 - \beta_1 - \beta_2 - \cdots \beta_p)$, when Y_t unit root, $\beta_1 + \beta_2 + \cdots \beta_p = 1$
  - using Yule-Walker Equation, ACF decays exponentially, PACF=0 after lag one

*AR(2): step1: 特征化等式: $Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} = \varepsilon_t$; step2: 特征根 $f(x) = 1 - \beta_1 x - \beta_2 x^2 = 0$; step3: solution

1. $\rho(h) = C_1\lambda_1^{-h} + C_2\lambda_2^{-h}$
2. $\rho(h) = \lambda^{-h}(C_1 + C_2 h)$
3. $\rho(h) = a|\lambda|^{-h}\cos(h\theta + b)$

$\lambda_1 = a + bi$, $\lambda_2 = a - bi$, $\tan\theta = \frac{b}{a}$ (这部分不考, 不用写2(h))

- moving average process of order q: $Y_t = C + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots \theta_q\varepsilon_{t-q}$, $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$
  - MA(1): $Y_t = C + \varepsilon_t + \theta\varepsilon_{t-1}$, ACF cuts off at $q=1$, PACF decays exponentially:
    $Y_t = C + (1 - (-\theta)L)\varepsilon_t$, $\varepsilon_t = \frac{-C}{1+\theta} + \frac{Y_t}{1+\theta L} = \frac{-C}{1+\theta} + Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} \cdots$
    $\therefore Y_t = \frac{C}{1+\theta} + \varepsilon_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \cdots$ ($|\theta| < 1$ for invertibility)
    $\therefore$ MA(1) ~ AR(∞) $\Rightarrow$ decay exponentially; PACF(h) $= \frac{\theta^h(1-\theta^2)}{1-\theta^{2(h+1)}}$
- ARMA(p, q): $B(L)Y_t = \mu + \Theta(L)\varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$
  - ARMA(1,1): $(1 - \beta_1 L)Y_t = \beta_0 + (1 + \theta_1 L)\varepsilon_t \Rightarrow Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1}$
    ACF and PACF both decays exponentially
  - Generally: $B(L) = (1 - \lambda_1 L)(1 - \lambda_2 L)\cdots(1 - \lambda_p L)$ with $|\lambda_i| < 1 \Rightarrow$ stationary
    $Y_t = \mu + \frac{\Theta(L)}{B(L)}\varepsilon_t = \mu + \Psi(L)\varepsilon_t$ ($\mu = \frac{\beta_0}{B(1)}$), $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$ (infinit lag polynomial)
- ARIMA(p, d, q): d = order of integration
  $d = 0 \Rightarrow$ ARMA(p, q), $d = 1 \Rightarrow \Delta Y_t \sim$ ARIMA(p-1, 0, q) = ARMA(p-1, q)

6 Unit Root Test

- unit root in AR(p) model: AR(p): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + U_t$
  $\Rightarrow \Delta Y_t = \beta_0 + \delta Y_{t-1} + r_1\Delta Y_{t-1} + r_2\Delta Y_{t-2} + \cdots r_{p-1}\Delta Y_{t-p+1} + U_t$ (subtract Y_{t-1})
  (where $\delta = \beta_1 + \beta_2 + \cdots \beta_p - 1$, $r_1 = -(\beta_2 + \cdots \beta_p)$, $r_2 = -(\beta_3 + \cdots \beta_p)$, $\cdots$, $r_{p-1} = -\beta_p$)
  (AR(2): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t \Rightarrow \Delta Y_t = \beta_0 + \delta Y_{t-1} + r_1\Delta Y_{t-1} + U_t$)

H_0: $\delta = 0$ (have unit root) $\Rightarrow$ AR(p) becomes AR(p-1) model in first difference.

(※) AR coefficients can be badly biased towards zero (if there is a unit root)

The distribution of the OLS estimator of the AR(1) coefficient: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$

$\hat\beta_1 = \frac{\sum_{t=1}^{T}(Y_{t-1} - \bar Y)Y_t}{\sum_{t=1}^{T}(Y_{t-1} - \bar Y)^2} = \beta_1 + \frac{\sum_{t=1}^{T}Y_{t-1}U_t}{\sum_{t=1}^{T}Y_{t-1}^2}$ (assume $\bar Y = Y_0 = 0$)

- ∘ when $|\beta_1| < 1$, $Var(\hat\beta_1) = \sigma_u^2/SST_{Y_{t-1}} \xrightarrow{p} \sigma_u^2/(T\sigma_Y^2) = \frac{1}{T}\sigma_u^2/\sigma_Y^2$, where $\sigma_Y^2 = \sigma_u^2/(1-\rho^2)$, $\Rightarrow Var(\hat\beta_1) \xrightarrow{p} \frac{1-\rho^2}{T}$
  Thus $\hat\beta_1 - \rho \sim N(0, (1-\rho^2)/T)$ (given that TS.1*-TS.5* are satisfied)
- ∘ when $\rho = 1$, $\hat\beta_1 - \rho = \frac{\sum_{t=1}^{T}Y_{t-1}U_t}{\sum_{t=1}^{T}Y_{t-1}^2}$ (下面的证明所示, 原因不讲) $\Rightarrow T(\hat\rho - \rho) = \frac{\frac{1}{T}\sum_{t=1}^{T}Y_{t-1}U_t}{\frac{1}{T^2}\sum_{t=1}^{T}Y_{t-1}^2}$
  - $\begin{cases} Y_t = Y_{t-1} + U_t \\ Y_t^2 = Y_{t-1}^2 + 2U_t Y_{t-1} + U_t^2 \end{cases}$, where $Y_T = Y_0 + (U_1 + U_2 + \cdots U_T) = \sum_{t=1}^{T}U_t$ (assume $Y_0 = 0$)
  - $\Rightarrow Y_{t-1}U_t = \frac{1}{2}(Y_t^2 - Y_{t-1}^2 - U_t^2)$; $\frac{1}{\sqrt T}Y_T \sim \frac{1}{\sqrt T}\sum_{t=1}^{T}U_t \xrightarrow{d} N(0, \sigma_u^2) = \sigma_u N(0,1)$
  - $\Rightarrow \sum Y_{t-1}U_t = \frac{1}{2}(Y_T^2 - Y_0^2 - \sum_{t=1}^{T}U_t^2)$; Thus, $Y_T^2/T \xrightarrow{d} \sigma_u^2\chi^2(1)$
  - $\therefore$ 分子 $= \frac{1}{T}\sum Y_{t-1}U_t = \frac{1}{2}(Y_T^2/T - 0 - \frac{\sum_{t=1}^{T}U_t^2}{T}) \xrightarrow{d} \frac{\sigma_u^2}{2}(\chi^2(1) - 1)$
  - 分母 $\xrightarrow{d} \int_0^1 B^2(r)dr$, 因此可以推导出 $T(\hat\rho - \rho)$ 的 distribution)

Dickey-Fuller Test: $Y_t = \alpha + \rho Y_{t-1} + e_t$ (ADF test: control for serial correlation)

H_0: $\rho = 1$ (the process has a unit root), H_1: $\rho < 1$ — under the Null, Y_t is nonstationary, t statistics has a DF distribution! (≠ Normal) not t distribution!

- Alternative: $\Delta Y_t = \alpha + \theta Y_{t-1} + e_t$ (subtract Y_{t-1})
  H_0: $\theta = 0$ (unit root), H_1: $\theta < 0$, critical values: -3.43 (1%), -3.12 (2.5%), -2.86 (5%), -2.57 (10%)
- Augmented Dickey-Fuller test: $\Delta Y_t = \alpha + \theta Y_{t-1} + r_1\Delta Y_{t-1} + \cdots r_p\Delta Y_{t-p} + e_t$
- with time trends: $\Delta Y_t = \alpha + \delta t + \theta Y_{t-1} + r_1\Delta Y_{t-1} + \cdots r_p\Delta Y_{t-p} + e_t$ (H_1: the process is trend stationary) — -3.41 (5%)

(*drift: $\alpha \neq 0$, $\delta \neq 0$ (constant / time trend))

7 Spurious Regression and Cointegration

Spurious Regression: reg on I(1)-series on another I(1)-series $\Rightarrow$ high t-statistics, high R^2 (maybe)

Cointegration: two I(1)-series Y_t, X_t are said to be cointegrated if:

$Y_t = \alpha + \beta X_t + \varepsilon_t$, with $\varepsilon_t \sim I(0)$ The deviation is quite stable

- EG-ADF test: test for cointegration if the parameters are unknown: test DF for $\hat u_t$ ($\hat\varepsilon_t$) — -3.34 (5%)
  including time trend: critical value $= -3.78$ (5%)
  (if: $\hat u_t$ has a unit root is rejected $\Rightarrow$ cointegration; superconsistency (endogeneity doesn't cause any inconsistency))
- Dynamic OLS (DOLS) estimator: making X_t strictly exogenous, use HAC s.e. $\Rightarrow$ serial correlation
- Error Correction Model: $\Delta Y_t = \alpha_0 + r_0\Delta X_t + \alpha(Y_{t-1} - \beta X_{t-1}) + U_t$ (ECM)

8 Forecasting

- Bayes Information Criterion $BIC(p) = \ln(\frac{SSR(p)}{T}) + (p+1)\frac{\ln T}{T} \Rightarrow \hat p^{BIC} \xrightarrow{p} p$
- Akaike Information Criterion $AIC(p) = \ln(\frac{SSR(p)}{T}) + (p+1)\frac{2}{T}$ — $2 < \ln T$, smaller penality, overestimate p!

1. AR(p): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_p Y_{t-p} + \varepsilon_t$ → first demean the equation
   $(Y_t - \mu) = \beta_1(Y_{t-1} - \mu) + \beta_2(Y_{t-2} - \mu) + \cdots + \beta_p(Y_{t-p} - \mu) + \varepsilon_t$ → thus the expectation comes to zero
   $\tilde Y_t = \beta_1\tilde Y_{t-1} + \beta_2\tilde Y_{t-2} + \cdots \beta_p\tilde Y_{t-p} + \varepsilon_t$, $E(\tilde Y_t) = 0$,
   multiplying throughout by $\tilde Y_t, \tilde Y_{t-1}, \cdots \tilde Y_{t-p}$ in turn and take expectations
   $Var(\tilde Y_t) = \beta_1\rho_1 + \beta_2\rho_2 + \cdots \beta_p\rho_p + \sigma^2$
   $\rho_1 = \beta_1 Var(\tilde Y_t) + \beta_2\rho_1 + \cdots \beta_p\rho_{p-1}$ } Yule-Walker Equations
   ACF decays exponentially, PACF=0 after lag p

   (ARMA(1,1): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1}$, stationary if $|\beta_1| < 1$, $E(Y_t) = \beta_0/(1-\beta_1)$, $Var(Y_t) = \sigma^2(1+\beta_1^2)/(1-\beta_1^2)$, 同样可以求 $\rho_1, \rho_2 \ldots$)

2. MA(q): $Y_t = C + \varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \cdots \theta^q\varepsilon_{t-q}$
   ACF cuts off at q, PACF decays exponentially.

Panel Data Models - Random Effect (supplement notes).

model: $Y_{it} = \beta_0 + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \cdots \beta_k X_{k,it} + a_i + U_{it}$, $V_{it} = a_i + U_{it}$, assumption: $Cov(X_{j,it}, a_i) = 0$

$Cov(V_{it}, V_{is}) = Cov(a_i, a_i) = \sigma_a^2$, $Corr(V_{it}, V_{is}) = \sigma_a^2/(\sigma_a^2 + \sigma_u^2)$, $\lambda = 1 - \sqrt{\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)}$

$(y_{it} - \lambda\bar y_i) = \beta_1(X_{1,it} - \bar X_{1i}) + \cdots \beta_k(X_{k,it} - \bar X_{ki}) + (V_{it} - \bar V_i)$

proof for $E(e_{it}e_{is}) = 0$, where $e_{it} = V_{it} - \lambda\bar V_i$ (下面用θ代替λ).

$E(e_{it}) = E(V_{it} - \lambda\bar V_i) = E(V_i) - \lambda E(\bar V_i) = 0$ — where $E(V_{it}\bar V_i) = \sigma_a^2 + \frac{1}{T}\sigma_u^2 = E(\bar V^2)$

$Var(e_{it}) = Var(V_{it} - \theta\bar V_i) = Var(V_{it}) + \theta^2 Var(\bar V_i) - 2\theta Cov(V_{it}, \bar V_i) = \sigma_V^2 + \theta^2 E(\bar V_i^2) - 2\theta E(V_{it}\bar V_i)$
$= (\sigma_a^2 + \sigma_u^2) + (-2\sqrt{\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)} + \sigma_u^2/(\sigma_u^2 + T\sigma_a^2)) + 2\sqrt{\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)})(\sigma_a^2 + \frac{1}{T}\sigma_u^2) = \sigma_u^2$

$E(e_{it}e_{is}) = E[(V_{it} - \theta\bar V_i)(V_{is} - \theta\bar V_{is})] = E(V_{it}V_{is}) - 2\theta E(\bar V_i V_{it}) + \theta^2 E(\bar V_i^2) = \sigma_a^2 - 2\theta(\sigma_a^2 + \frac{1}{T}\sigma_u^2) + \theta^2(\sigma_a^2 + \frac{1}{T}\sigma_u^2) = 0$

- estimation of $\theta = 1 - \sqrt{\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)}$: using $\hat\sigma_u^2$ and $\hat\sigma_a^2$
  $\hat\sigma_a^2 = \frac{1}{nT(T-1)/2 - (k+1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat V_{it}\hat V_{is}$, $\hat V_{it}$ is the residual of pooled OLS
  $\hat\sigma_V^2 = \sum_{t=1}^{T}\hat V_{it}^2/(nT - k)$, thus $\hat\sigma_u^2 = \hat\sigma_V^2 - \hat\sigma_a^2$ could be estimated
  $\Rightarrow$ estimate $\hat\theta$ (consistent, not unbiased, asymptotically normal distributed)
  (proof: $\sum_{t=1}^{T}\sum_{s=t+1}^{T}E(V_{it}V_{is}) = \frac{T(T-1)}{2}\sigma_a^2$ consider the following right-top block

$$Var(V_i|X_i) = \begin{pmatrix} \sigma_a^2 + \sigma_u^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_u^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 + \sigma_u^2 \end{pmatrix} \Rightarrow E(V_{i1}V_{i2}) + E(V_{i1}, V_{i3}) + \cdots E(V_{i1}, V_{iT}) + E(V_{i2}, V_{i3}) + \cdots E(V_{i2}, V_{iT}) \cdots E(V_{iT-1}, V_{iT}))$$

Chapter1 Review for Econometrics1

unbiasedness: $E(u|X)=0$
asy.consistency: $E(u)=0, Cov(u,X)=0$

Linear Regression Model: $Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots\beta_kX_{ik}+u_i$. $R^2=1-\frac{SSR}{SST}=1-\frac{\hat\sigma_u^2}{\hat\sigma_Y^2}$

OLS estimators: $\hat\beta_j=\frac{\text{sample }Cov(Y,X_j)}{\text{sample }Var(X_j)}$, or $\hat\beta=(X'X)^{-1}X'Y=(\sum_{i=1}^n X_iX_i')^{-1}\sum_{i=1}^n X_iY_i$

- SLR: $\hat\beta_1=\frac{\sum_{i=1}^n(X_i-\bar X)Y_i}{\sum_{i=1}^n(X_i-\bar X)^2}=\beta_1+\frac{\sum_{i=1}^n(X_i-\bar X)u_i}{\sum_{i=1}^n(X_i-\bar X)^2}$ • MLR: $\hat\beta_1=\frac{\sum_{i=1}^n\hat r_{i1}Y_i}{\sum_{i=1}^n\hat r_{i1}^2}=\beta_1+\frac{\sum\hat r_{i1}u_i}{\sum\hat r_{i1}^2}$ (reg X_{i1} on X_j (X_{-1})) R_j: reg X_j on other regressors
- SLR: $Var(\hat\beta_1)=\frac{\sigma^2}{SST_x}$, $Var(\hat\beta_0)=\sigma^2(\frac{1}{n}+\frac{\bar X^2}{SST_x})=\frac{\sigma^2\sum X_i^2}{nSST_x}$ • MLR: $Var(\hat\beta_j)=\frac{\sigma^2}{SST_j(1-R_j^2)}=\frac{\sigma^2}{\sum\hat r_{ij}^2}$, $VIF=\frac{1}{1-R_j^2}$

$Var(\hat\beta_j)=\frac{\sum(\hat r_{ij}\hat u_i)^2}{SST_x^2}$, $se(\hat\beta_j)=\sqrt{\frac{\sum(\hat r_{ij}\hat u_i)^2}{SST_x^2}}$ (SLR)

$Var(\hat\beta_j)^2=\frac{\sum\hat r_{ij}^2\hat u_i^2}{SSR_j^2}$, $se(\hat\beta_j)^2=\frac{\sum\hat r_{ij}^2\hat u_i^2}{SSR_j^2}$ (MLR), $(SSR_j=\sum\hat r_{ij}^2)$

Estimating σ^2: $\hat\sigma^2=SSR/(n-k-1)$; adjusted R^2: $\bar R^2=1-\frac{SSR/(n-k-1)}{SST/(n-1)}$, $se(\hat\beta_j)=\frac{\hat\sigma}{\sqrt{SST_j(1-R_j^2)}}$

statistics: $z=\frac{\hat\beta_j-\beta_j}{sd(\hat\beta_j)}\sim N(0,1)$, $t=\frac{\hat\beta_j-\beta_j}{se(\hat\beta_j)}\sim t(n-k-1)$, $F=\frac{(SSR_r-SSR_{ur})/q}{SSR_{ur}/(n-k-1)}=\frac{(R^2_{ur}-R^2_r)/q}{(1-R^2_{ur})/(n-k-1)}\sim F(q,n-k-1)$ (restricted)

consistency: $\hat\beta_j\xrightarrow{p}\beta_j+\frac{Cov(u,X_j)}{Var(X_j)}\xrightarrow{p}\beta_j$, $z,t,F\xrightarrow{d}$ above distribution, $LM=nR^2_{\hat u}\xrightarrow{d}\chi^2(q)$ (reg $\hat u_i$ on $1,X_{i1},\cdots X_{ik}$)

Heteroskedasticity: $Var(u_i|X)=\sigma_i^2=\sigma^2(X_i)^2$, White's robust $Var(\hat\beta_j)$ and $se(\hat\beta_j)$

- Breusch-Pagan Test: $F=\frac{R^2_{\hat u^2}/k}{(1-R^2_{\hat u^2})/(n-k-1)}\sim F(k,n-k-1)$, $R^2_{\hat u^2}$ = reg $\hat u^2$ on X, $LM=nR^2_{\hat u^2}\xrightarrow{d}\chi^2(k)\Rightarrow$ Linear form of σ_i
- White Test: reg $\hat u^2$ on $(1,X,X^2,X_iX_j)$, $LM=nR^2_{\hat u^2}\xrightarrow{d}\chi^2(df)$; special case: reg on $(1,\hat Y_i,\hat Y_i^2)\Rightarrow nR^2_{\hat u^2}\xrightarrow{d}\chi^2(2)$

GLS is a special case of WLS (used when σ_i is known) ⇒ BLUE

FGLS (used when conditional variance is unknown) ⇒ run $(\log\hat u^2$ on X), estimate $\hat\sigma_i^2$ ⇒ not unbiased, consistent. Even when the assumed function is wrong, more efficient than OLS

Chapter 2 Endogeneity

definition: $Cov(X_i,u_i)\neq0$ — $E(u|X)=0$: mean independence (X stochastic) ⇒ 1. $E(u)=0$, 2. $Cov(u,X)=0$, 3. $E(u|X_i)=0$

- OVB: $\hat\beta_1\xrightarrow{p}\beta_1+Cov(X,u)/Var(X)$, $Y=\beta_0+\beta_1X_1+(\beta_2X_2+u)$, in this case, the OVB is $\beta_2Cov(X_1,X_2)/Var(X_1)$ for $\hat\beta_1$

  solutions: 1. include more regressors 2. use panel data 3. use IV 4. run randomized controlled experiment
- Wrong Functional Form: $Y=\beta_0+\beta_1X_1+(\beta_2X_1^2+u)$

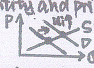
  solutions: 1. continous dependent variable, use nonlinear X 2. discrete/binary: probit/logic analysis
- Measurement Error: (true model) $Y=\beta_0+\beta_1Z+V$, (Measurement) $X=Z+W$

  $\Rightarrow Y=\beta_0+\beta_1(X-W)+V\Rightarrow Y=\beta_0+\beta_1X+(V-\beta_1W)$ — attenuation bias

  classical assumptions: $Cov(Z,W)=0$, thus $Cov(X,V-\beta_1W)\neq0$

  $\Rightarrow Cov(X,V-\beta_1W)=-\beta_1Var(W)=-\beta_1\sigma_W^2$, $\therefore\hat\beta_1\xrightarrow{p}\beta_1+Cov(X,V-\beta_1W)/Var(X)=\beta_1-\beta_1\frac{\sigma_W^2}{\sigma_X^2}=\frac{\sigma_Z^2}{\sigma_Z^2+\sigma_W^2}\beta_1$
- Simultaneity (quantity and price example)

  $\begin{cases}Q_i^D=a_1+b_1P_i+u_i\\Q_i^S=a_2+b_2P_i+V_i\end{cases}$ $(b_1<0,\ b_2>0)$



  Structural Model: $Q_i^D=Q_i^S$, $P_i^D=P_i^S)\Rightarrow Q=Q(u,V)$, $P=P(u,V)$

  $Cov(P,u)>0$, $Cov(P,V)<0$, using the fact that: (fitted line)

  $\Rightarrow\hat b_1=Cov(P,Q)/Var(P)$, $\hat a_1=E(Q)-E(P)\cdot Cov(P,Q)/Var(P)$

  Solutions for Endogeneity: IV

$Var(\hat u_i)=Var(u_i-\cdots)$; $Corr(u_{it},\Delta u_{it})=-0.5$; $\hat\sigma_u=\frac{SSR}{n-k}$; $\hat\sigma_a^2=\frac{\cdots}{nT(T-1)/2-k}$

Chapter 3 Instrumental Variables

requirements: 1. instrumental relevance $Cov(Z,X)\neq0$, 2. instrumental exogeneity $Cov(Z,u)=0$

- Model: $Y_i=\beta_0+\beta_1X_i+u_i$, $Cov(X_i,u_i)\neq0$, Z_i is an IV for X_i ⇒ unbiasedness 需 $E(u|X,Z)=0$

  $\hat\beta_{1,IV}=\frac{\text{Sample }Cov(Y_i,Z_i)}{\text{Sample }Cov(X_i,Z_i)}=\frac{\sum_{i=1}^n(Y_i-\bar Y)(Z_i-\bar Z)}{\sum_{i=1}^n(X_i-\bar X)(Z_i-\bar Z)}=\beta_1+\sum_{i=1}^nE\left(\frac{(u_i-\bar u)(Z_i-\bar Z)}{\sum(X_i-\bar X)(Z_i-\bar Z)}\right)\xrightarrow{p}\beta_1+\frac{Cov(Z,u)}{Cov(Z,X)}=\beta_1$
- assymptotic normality: ($V_i=(Z_i-\bar Z)u_i$)

  $\sqrt n(\hat\beta_{1,IV}-\beta_1)=\frac{\frac{1}{\sqrt n}\sum(u_i-\bar u)(Z_i-\bar Z)}{\frac1n\sum(X_i-\bar X)(Z_i-\bar Z)}=\frac{\frac{1}{\sqrt n}\sum(Z_i-\bar Z)u_i}{\frac1n\sum(X_i-\bar X)(Z_i-\bar Z)}\xrightarrow{p}\frac{\frac{1}{\sqrt n}\sum V_i}{Cov(X,Z)}=\frac{\sigma_V}{Cov(Z,X)}\cdot\frac{\frac{1}{\sqrt n}\sum V_i}{\sigma_V}\xrightarrow{p}\frac{\sigma_V}{Cov(Z,X)}\cdot N(0,1)$

  $\xrightarrow{d}N\left(0,\frac{\sigma_V^2}{(Cov(Z,X))^2}\right)=N\left(0,\frac{\sigma_u^2Var(Z_i)}{[Cov(Z,X)]^2}\right)$, under MLR.5: $Var(u_i|Z_i)=\sigma_u^2$

  asymptotic variance $=\frac{\sigma_u^2}{nVar(X)}\cdot\frac{Var(X)Var(Z)}{Cov(Z,X)^2}=asym.Var(\hat\beta_{OLS})\cdot\frac{1}{\rho^2_{X,Z}}$ ⇒ estimate $\hat\sigma_u$: $\hat\sigma_u^2=\frac{\sum\hat u_i^2}{n-k}$
- insights: $Y_i=\beta_0+\beta_1X_i+u_i=\beta_0+\beta_1(\gamma_0+\gamma_1Z_i+V_i)+u_i=(\beta_0+\beta_1\gamma_0)+\beta_1\gamma_1Z_i+(\beta_1V_i+u_i)\Rightarrow\hat\beta_{IV}=\hat\beta\hat\gamma/\hat\gamma$

$se(\hat\rho)$ ($se(\hat\beta_{FD})$) (test for serial correlation, test whether $\Delta u_{it}=\rho\Delta u_{it-1}+v_i$, $\rho=-0.5$)

2SLS estimator: $X=Z\Gamma+U$, $Y=\hat X\beta+V$, where $\hat X=Z\hat\Gamma$, which is the clean part

- SLR: $\hat\beta_{2SLS}=\frac{\sum_{i=1}^n(Y_i-\bar Y)(\hat X_i-\bar{\hat X})}{\sum_{i=1}^n(\hat X_i-\bar{\hat X})^2}=\frac{\sum(Y_i-\bar Y)\hat\Gamma(Z_i-\bar Z)}{\sum(Z_i-\bar Z)^2\hat\Gamma^2}=\frac{\sum(Y_i-\bar Y)(Z_i-\bar Z)}{\sum(Z_i-\bar Z)^2\hat\Gamma}$, where $\hat\Gamma=\frac{\sum(Z_i-\bar Z)(X_i-\bar X)}{\sum(Z_i-\bar Z)^2}$

  $\Rightarrow\hat\beta_{2SLS}=\hat\beta_{IV}$ (when SLR) when SLR + m=1: $\hat\beta^{2SLS}=\hat\beta^{MOM}=\hat\beta_{IV}$
- MLR: $Y=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\cdots\beta_kX_{ki}+\beta_{k+1}W_{1i}+\cdots\beta_{k+r}W_{ri}+u_i$ — {k endogenous X_i; r exogenous W_i; m exogenous IV Z_i}
  1. (assume k=1). $X_i=\Pi_0+\Pi_1Z_{1i}+\Pi_2W_i+V_i\Rightarrow X_i=\hat X_i+\hat V_i$
  2. structural model: $Y_i=\beta_0+\beta_1\hat X_i+\beta_2W_i+e_i$ $(e_i=\beta_1\hat V_i+u_i)$, $Var(\hat\beta_1)=\frac{\sigma^2}{SST_{\hat X}(1-R^2_{\hat X})}$

  bigger m: reduce asymptotic variance but seems to increase finit sample bias

  $Cov(Z,u)\neq0$: $\begin{cases}plim\hat\beta_{IV}-\beta_1=\frac{Corr(Z,u)}{Corr(Z,X)}\cdot\frac{\sigma_u}{\sigma_X}\\plim\hat\beta_{OLS}-\beta_1=Corr(X,u)\cdot\frac{\sigma_u}{\sigma_X}\end{cases}$ $X_i=X_{1i}+X_{2i}$ (clean part + dirty part); $Z_{1i},\cdots Z_{mi}$ (m↑) ⇒ larger clean part ⇒ $SST_{\hat X}$↑

Checking Instrument Relevance: $Cov(Z,X)\neq0$ (deal with serial correlation)

(first stage regression) $X_i=\Pi_0+\Pi_1Z_{1i}+\cdots\Pi_mZ_{mi}+\Pi_{m+1}W_{1i}+\cdots\Pi_{m+r}W_{ri}+V_i$

$\begin{cases}H_0:\Pi_1=\Pi_2=\cdots\Pi_m=0\text{ or nearly zero (weak instruments) explain very little beyond }W_i\\H_1:\text{ one or more nonzeros}\end{cases}$ H_0: failure for t test even when n is large

Rule of thumb: F>10, reject the null ⇒ the bias of 2SLS relative to OLS is less than 10%

Test for Instrument Exogeneity: $Cov(Z_i,u_i)=0$ (J-test)

- cannot test when m=k: sample $Corr(Z_i,\hat u_i)=\sum_{i=1}^nZ_i\hat u_i=0$ (FOC2) → $\hat\beta_{2SLS}=(Z'X)^{-1}Z'Y$, $\hat u=Y-X\hat\beta^{2SLS}$ ⇒ $Z'\hat u=Z'Y-Z'X(Z'X)^{-1}Z'Y=0$
- can test when m>k
  1. obtain $\hat u_i^{2SLS}=Y_i-\hat\beta_0-\hat\beta_1X_i-\hat\beta_2W_i$ (where $\hat\beta$ are estimated by 2SLS)

     notice that $\hat u_i^{2SLS}=(Y_i-\hat\beta_0-\hat\beta_1\hat X_i-\hat\beta_2W_i)-\hat\beta_1\hat V_i$, given $X_i=\hat X_i+\hat V_i$

     when m=k, $\sum_{i=1}^nZ_i\hat u_i^{2nd}=0$ (FOC 2), dirty part $\sum_{i=1}^nZ_i\hat V_i=0$, thus $\sum_{i=1}^nZ_i\hat u_i=0$
  2. regression: $\hat u^{2SLS}=\delta_0+\delta_1Z_{1i}+\delta_2Z_{2i}+\cdots\delta_mZ_{mi}+\delta_{m+1}W_{1i}+\cdots\delta_{m+r}W_{ri}+e_i$

     $\begin{cases}H_0:\delta_1=\delta_2=\cdots\delta_m=0\ (Z\text{ exogenous})\\H_1:\text{ instrument endogeity}\end{cases}$ J-test: $J=mF\xrightarrow{asy}\chi^2(m-k)$, where $F=\frac{(R^2_{ur}-R_r^2)/m}{(1-R^2_{ur})/(n-m-r-1)}$, J=0 when m=k

Test for Endogeneity: $Cov(X,u)\neq0$ (Hausman test)

structural equation: $Y=\beta_0+\beta_1X+\beta_2W+u_i$

reduced-form equation: $X=\Pi_0+\Pi_1Z_1+\Pi_2Z_2+\Pi_3W+V\Rightarrow\hat V_i=X_i-\hat\Pi_0-\hat\Pi_1Z_{1i}-\hat\Pi_2Z_{2i}-\hat\Pi_3W_i$ (first-stage residual)

control function regression: $Y_i=\beta_0+\beta_1X_i+\beta_2W_i+\delta\hat V_i+e_i$

$\begin{cases}H_0:\delta=0\ (\text{exogenous})\\H_1:\delta\neq0\ (\text{endogenous})\end{cases}$ *Multiple endogenous variables: $H_0:\delta_1=\delta_2=\cdots\delta_k=0$ (jointly test the residuals from the first stage)

- OLS estimator $\hat\beta$ here is consistent: $X=\hat X+\hat V$, dirty part included ⇒ $Cov(X_i,e_i)=0$
- OLS estimator $\hat\beta$ here is same as the 2SLS estimator

  denote $X_{i1}=[X_i\ W_i]'$, $X_{i2}=\hat V_i$ (the partialling-out process)
  1. reg X_{i1} on X_{i2}: $\hat u_{i1}=[\hat X_i\ W_i]'$
  2. reg Y_i on $\hat u_{i1}$: $\hat\beta_{OLS}$ = reg Y_i on $\hat X_i$ — the same as $\hat\beta^{2SLS}$

$y=x'Ax\Rightarrow\frac{\partial y}{\partial x}=(A+A')x$

Chapter4 Binary Dependent Variable

1. Linear Probability Model: $Y_i=\beta_0+\beta_1X_{1i}+\cdots\beta_kX_{ki}+u_i$ (Y_i is dummy variable, $E(u_i|X_{1i},\cdots X_{ki})=0$)
   - intepretation: $E(Y_i|X)=1\cdot Pr(Y=1|X)+0\cdot Pr(Y=0|X)=Pr(Y=1|X)$
   - heteroskedasticity: $Var(Y|X)=E(Y^2|X)-E(Y|X)^2=p(Y=1|X)-p^2(Y=1|X)=p(Y=1|X)[1-p(Y=1|X)]$ (as we have $p(X)=E(Y|X)=Pr(Y=1|X)$, $Var(Y|X)=p(X)(1-p(X))$)

     where $\hat p(X)=\hat E(Y|X)=\hat\beta_0+\hat\beta_1X$ — binary errors: $u_i|X_{1i},\cdots X_{ki}=1-p_i$ (with prob. p_i); $-p_i$ (with prob. $1-p_i$)
   - disadvantage: 1. predicted probability <0 or >1 2. probability linear in X's assumption may error
2. Probit and Logit Model: $Y_i=G(\beta_0+\beta_1X_{1i}+\cdots\beta_kX_{ik})+u_i$
   - probit model: standard normal CDF, $Pr(Y=1|X)=\Phi(z)=\Phi(\beta_0+\beta_1X)$, where $\Phi(z)=\int_{-\infty}^z\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$
   - logit model: logistic CDF: $Pr(Y=1|X)=\Lambda(z)=\frac{exp(z)}{1+exp(z)}=\frac{exp(\beta_0+X\beta)}{1+exp(\beta_0+X\beta)}$ → similiar predications

   (1) Nonlinear Least Squares (NLS)

   $(\hat\beta_0,\hat\beta_1,\cdots\hat\beta_k)=\arg\min_{(\beta)}\sum_{i=1}^n[Y_i-G(\beta_0+\beta_1X_{i1}+\cdots\beta_kX_{ik})]^2$ — consistent, asymptotically normally d., inefficient

   (2) Maximum Likelihood Estimation (MLE)

   conditional probability distribution for the ith observation: $P(Y_i=y_i|X_i)=p_i^{y_i}(1-p_i)^{1-y_i}$

   joint probability distribution of $Y_1,\cdots Y_n$ conditional on X's is: (assume (X_i,Y_i) i.i.d.)

   $P(Y_1=y_1,\cdots Y_n=y_n|X_1,\cdots X_n)=\prod_{i=1}^nP(Y=y_i|X_i)=\prod_{i=1}^np_i^{y_i}(1-p_i)^{1-y_i}$

   $\Rightarrow L(\beta_0,\beta_1,\cdots\beta_k)=\ln\prod_{i=1}^np_i^{y_i}(1-p_i)^{1-y_i}=\sum_{i=1}^n\{y_i\ln G(\beta_0+\beta_1X_{i1}+\cdots\beta_kX_{ik})+(1-y_i)\ln[1-G(\beta_0+\beta_1X_{i1}+\cdots\beta_kX_{ik})]\}$

   → assistent, asymptotically normally distributed, efficient, commonly used

   ⇒ sign and significance are comparable between the two models
   - PEA (Partial Effect at the Average): $g(\hat\beta_0+\hat\beta_1\bar X_1+\cdots\hat\beta_k\bar X_k)\hat\beta_j$: problematic when Y_i = gender (e.g.)
   - APE (Average Partial Effect), $\frac1n\sum_{i=1}^ng(\hat\beta_0+\hat\beta_1X_{i1}+\cdots\hat\beta_kX_{ik})\hat\beta_j$: make more sense
3. Good of fit measures: percent correctly predicted (correctly predictly 1s, 0s, all)

   pseudo R-squared: $\tilde R^2=1-L_{ur}/L_0$, $\frac{L_{ur}}{L_0}=\frac{|L_{ur}|}{|L_0|}$, $|L_{ur}|\le|L_0|$

   correlation based measures: $Corr(Y_i,\hat Y_i)^2$

   $LR=2(L_{ur}-L_r)\sim\chi^2(q)$, H_0: $L_{ur}-L_r$ closed to zero (L_r L_{ur} 0)

Chapter5 Panel Data

{pooled cross-section: sampling randomly from a large population at different points in time; panel data: the same cross-section units over time (same set of units)}

- Pooled Cross-Section: 1. bigger sample size 2. investigate effect of time 3. changing relations over time
  1. Dummy Variables Approach / Difference-in-Difference estimators (DID)

     $Y=\beta_0+\delta_0d+\beta_1X_1+\delta_1d\cdot X_1+\cdots\beta_kX_k+\delta_kd\cdot X_k+u$, DID estimators: $\delta_1,\delta_2,\cdots\delta_k$

     $\begin{cases}H_0:\text{no change over time }(\delta_0=\delta_1=\cdots\delta_k=0)\\H_1:\text{relationships have changed.}\end{cases}$ restricted model: $Y=\beta_0+\beta_1X_1+\cdots\beta_kX_k+u$

     $\Rightarrow F=\frac{(SSR_r-SSR_{ur})/(k+1)}{SSR_{ur}/[n-2(k+1)]}=\frac{(R^2_{ur}-R^2_r)/(k+1)}{(1-R^2_{ur})/[n-2(k+1)]}\sim F(k+1,n-2(k+1))$

     policy analysis - Natural/Quasi Experiment

	Before		After		
β_0	control group	exogenous event (policy change) →	control group	$\beta_0+\delta_0$	no effect of policy
$\beta_0+\beta_1$	treatment group		treatment group	$\beta_0+\delta_0+\beta_1+\delta_1$	affected by policy

treatment effect: $\hat\delta_1=(\bar Y_{T,after}-\bar Y_{T,before})-(\bar Y_{C,after}-\bar Y_{C,before})$ — trends assumption

2. Chow Test (restriction: normality)

   restricted models: $\begin{cases}Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots\beta_kX_k+u & (n_1\text{ observations})\\Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots\beta_kX_k+u & (n_2\text{ observations})\end{cases}$ (n_1+n_2 observations)

   unrestricted models: $Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots\beta_kX_k+u$

   H_0: $\beta_0=\beta_0'$ ($j=0,1,\ldots k$) no changed relationship; H_1: relationship has changed $\Rightarrow F=\frac{[SSR-(SSR_1+SSR_2)]/(k+1)}{(SSR_1+SSR_2)/[n-2(k+1)]}\sim F(k+1,n-2(k+1))$

- Panel Data Analysis (longitudinal data)

	time period 1	time period 2	⋯ T periods (t=1,2,3,⋯T)
entities (i=1,2,⋯n)	(Y_{11},X_{11})	⋯	

   balanced panel: no missing observations

1. Two-Period Panel Data Analysis (DID): $Y_{it}=\beta_0+\delta_0d2_t+\beta_1X_{it}+a_i+u_{it}$ — a_i: unobserved fixed effect (heterogeneity); u_{it}: idiosyncratic time-varying error

   also known as First-Difference method: $\Delta Y_i=\delta_0+\beta_1\Delta X_i+\Delta u_i$ — eliminate time-invariant endogeneity $Cov(X_i,a_i)\neq0$

   conditions: 1. $E(\Delta u_i)=0$, 2. $E(\Delta u_i\Delta X_i)=0$, 3. $E(\Delta u_i|X_i,a_i)=0$ (strict exogeneity)
2. Fixed-Effect Estimation (LSDV)

   $Y_{it}=\beta_0+\beta_1X_{it1}+\beta_2X_{it2}+\cdots\beta_kX_{itk}+a_i+u_{it}$, $t=1,2,\cdots n$, $t=1,2,\cdots T$

   $\Rightarrow\bar Y_i=\beta_1\bar X_{i1}+\beta_2\bar X_{i2}+\cdots\beta_k\bar X_{ik}+\bar a_i+\bar u_i$, where $\bar Y_i=\frac1T\sum_{t=1}^TY_{it}$

   $\Rightarrow(Y_{it}-\bar Y_i)=\beta_1(X_{it1}-\bar X_{i1})+\beta_2(X_{it2}-\bar X_{i2})+\cdots\beta_k(X_{itk}-\bar X_{ik})+(u_{it}-\bar u_i)$

   or: $\ddot Y_{it}=\beta_1\ddot X_{it1}+\cdots\beta_k\ddot X_{itk}+\ddot u_{it}$ (time-demeaned equation)

   pooled OLS: fixed effect / within estimators

   OLS on $Y_{it}=\beta_0+\beta_1X_{it}+a_i+u_{it}$ ⇒ between estimators (biased when $Cov(X_i,a_i)\neq0$)

   conditions | restrictions: 1. strict exogeneity 2. $df=NT-N-K$

   interpretation (dummy variable approach): $Y_{it}=a_1d1_i+a_2d2_i+\cdots a_Nd_{Ni}+\beta_1X_{it1}+\cdots+u_{it}$

   estimate individual effect: $\hat a_i=\bar Y_i-\hat\beta_1\bar X_{i1}-\cdots\hat\beta_k\bar X_{ik}$, $i=1,\cdots N$
   - When T=2, FE = FD
   - When T≥3, FE ≠ FD: FE better when u_{it} are serially uncorrelated (especially white noise $u_{it}=u_{it-1}+\varepsilon_{it}$); FD better: u_{it} severely serial-correlated, T→N or T>N

   Asy.FE1. $Y_{it}=\beta_1X_{it1}+\cdots\beta_kX_{itk}+a_i+u_{it}$, where a_i is unobserved

   Asy.FE2. random sampling across section, $(X_{i1},\cdots X_{ik},u_i,a_i)$ iid

   Asy.FE3. each X_j changes over time; no perfect collinearity among X_j

   Asy.FE4. $E(u_{it}|X_i,a_i)=0$ strict exogeneity ⇒ FE estimator unbiased, FD: same

   Asy.FE5. $Var(u_{it}|X_i,a_i)=Var(u_{it})=\sigma_u^2$ ⇒ FD.5 $Var(\Delta u_{it}|X_i)=\sigma^2$ no serial correlation

   Asy.FE6. $Cov(u_{it},u_{is}|X_i,a_i)=0$; FD.6: $Cov(\Delta u_{it},\Delta u_{is}|X_i)=0$ ⇒ BLUE with [illegible]

   Asy.FE7. $u_{it}|X_i,a_i\sim N(0,\sigma_u^2)$ implies FE4-6 ⇒ exact t/F distribution
3. Random Effect Models

   Asy RE1. no perfect multicolinearity among X_j

   Asy RE2. $E(a_i|X_i)=\beta_0$ = const. in addition to FE4

   Asy RE3. $Var(a_i|X_i)=\sigma_a^2$, in addition to FE5

   composite error uncorrelated with explanatory X but is serially correlated

   Random Effect assumptions: $Cov(X_{itj},a_i)=0$, $j=1,2,\cdots k$ (RE.P) $Cov(v_{it},v_{is})=\sigma_a^2$

   $Y_{it}=\beta_0+\beta_1X_{it1}+\beta_2X_{it2}+\cdots\beta_kX_{itk}+a_i+u_{it}$, $v_{it}=a_i+u_{it}$ ⇒ consistent

   Quasi-demeaned: $(Y_{it}-\lambda\bar Y_i)=\beta_1(X_{it1}-\lambda\bar X_{i1})+\cdots\beta_k(X_{itk}-\lambda\bar X_{ik})$, with $\lambda=1-\sqrt{\frac{\sigma_u^2}{\sigma_u^2+T\sigma_a^2}}$

   (use FGLS to estimate λ) ⇒ consistent (not unbiased), asymptotically normally distributed; $[a_i-\lambda a_i+u_{it}-\lambda\bar u_i]$ satisfy GM assumption
   - RE better than POLS: removes some serial correlation: $a_i-\lambda a_i$ in error term
   - RE better than FE: can estimate time-invariant variables

   Correlated Random Effect Approach: $a_i=\alpha+\gamma_1\bar X_{i1}+\gamma_2\bar X_{i2}+\cdots\gamma_k\bar X_{ik}+r_i$ (CRE)

   $\Rightarrow Y_{it}=\beta_0+\alpha+\beta_1X_{it1}+\gamma_1\bar X_{i1}+\cdots\beta_kX_{itk}+\gamma_k\bar X_{ik}+[r_i+u_{it}]$ — estimators similar to FE

   Hausman Test: FE or RE? Are they different? (time-invariant vars included)

   $(\hat\beta_{RE}-\hat\beta_{FE})'Var(\hat\beta_{RE}-\hat\beta_{FE})^{-1}(\hat\beta_{RE}-\hat\beta_{FE})\xrightarrow{d}\chi^2(k)$ H_0: $Cov(a_i,X_{it})=0$

   H_0 under: $\hat\beta_{RE}$ and $\hat\beta_{FE}$ are both consistent, and $\hat\beta_{RE}$ is more efficient.

   H_1: only $\hat\beta_{FE}$ is consistent, $\hat\beta_{RE}$ is not consistent. $Cov(a_i,X_{it})\neq0$.
4. Clustered Standard Errors

   mean estimator: $\bar Y=\frac1n\sum_{i=1}^n\bar Y_i\xrightarrow{d}N(0,\frac{\bar\sigma^2}{n})$, where $\bar\sigma^2=Var(\bar Y_i)=Var(\frac1T\sum_{t=1}^TY_{it})$

   clustered SE of $\bar Y=\sqrt{s^2/n}$, where $s^2=\frac{1}{n-1}\sum_{i=1}^n(\bar Y_i-\bar Y)^2$

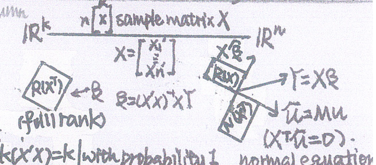
   $\Rightarrow s^2=\frac{1}{n-1}\sum_{i=1}^n[\frac1T\sum_{t=1}^T(Y_{it}-\bar Y)]^2=\frac{1}{n-1}\sum_{i=1}^n\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T(Y_{it}-\bar Y)(Y_{is}-\bar Y)$, estimating $\bar\sigma^2$ — positive autocorrelation; underestimate $\bar\sigma^2$

   clustered SEs are valid whether or not there is heteroskedasticity/serial correlation

Matrix Algebra

Regression Matrix Form: $Y=X\beta+U$ (add "1" to the first column)

$$\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix}=\begin{bmatrix}1&X_{11}&\cdots&X_{1k}\\1&X_{21}&\cdots&X_{2k}\\\vdots&&&\vdots\\1&X_{n1}&\cdots&X_{nk}\end{bmatrix}_{n\times k}\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_k\end{bmatrix}+\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}$$



Assumption 1. (Linear in β): $Y_i=X_i'\beta+u_i$

Assumption 2. (Nonsingularity): $rank(X'X)=k$ with probability 1

Assumption 3. (Strict Exogeneity): $E(\varepsilon_i|X)=E(\varepsilon_i|X_1,\cdots X_n)=0$

$\Rightarrow E(\varepsilon_i)=0$, $E(X_{jk}\varepsilon_i)=0$, if (Y_i,X_i) randomly chosen, $E(\varepsilon_i|X)=E(\varepsilon_i|X_i)=0$

Assumption 4. (Spherical error variance): $E(\varepsilon_i^2|X)=\sigma^2>0$, $E(\varepsilon_i\varepsilon_j|X)=0$ $(i\neq j)$

⇒ conditional homoskedasticity + conditional serial uncorrelation

$\Rightarrow Var(\varepsilon_i)=\sigma^2$, $Cov(\varepsilon_i,\varepsilon_j)=0$ and $Var(\varepsilon|X)=E(\varepsilon\varepsilon'|X)=\sigma^2I_n$

Asy.3 + Asy.4 ⇒ $E(\varepsilon|X)=0$ and $Var(\varepsilon|X)=E(\varepsilon\varepsilon'|X)=\sigma^2I_n$

OLS: $\hat\beta=\hat\beta_{OLS}=\arg\min_{\beta\in R^k}SSR(\beta)$, $SSR=(Y-X\beta)'(Y-X\beta)=Y'Y-2\beta'X'Y+\beta'X'X\beta$

⇒ FOC: $\frac{\partial SSR(\beta)}{\partial\beta}=-2X'Y+2X'X\beta=0\Rightarrow\hat\beta=(X'X)^{-1}X'Y=(\sum X_iX_i')^{-1}(\sum X_iY_i)$

SOC: $\frac{\partial^2SSR(\beta)}{\partial\beta\partial\beta'}=2X'X$: positive definite, check global minimum ✓

⇒ when an intercept is included: $X_{i1}=1$, $\sum\hat\varepsilon_i=0$

- estimation of $\sigma^2=E(\varepsilon_i^2)$, $\hat\sigma^2=\frac1n\sum\hat\varepsilon_i^2$ (MLE), unbiased $s^2=\frac{1}{n-k}\sum\hat\varepsilon_i^2=\frac{SSR}{n-k}$

Projection model: $P=X(X'X)^{-1}X'$ · symmetric · idempotent ($P^2=P$) · $PX=X$ (also symmetric and idempotent)

related matrix: $M=I-P=I-X(X'X)^{-1}X'$ · $tr(P)=rank(P)=k$, $tr(M)=rank(M)=n-k$

$(I-P)Y=MY=M(X\beta+\varepsilon)=M\varepsilon$ · $MX=0$, $PM=0$

$SSR=\hat\varepsilon'\hat\varepsilon=\varepsilon'M'M\varepsilon=\varepsilon'M\varepsilon$

Variance decomposition: 1. $Y'Y=\hat Y'\hat Y+\hat\varepsilon'\hat\varepsilon$, $\sum Y_i^2=\sum\hat Y_i^2+\sum\hat\varepsilon_i^2$

($M^0=I_n-\frac1n\iota\iota'$, demeaned matrix × M^0, $\iota=[1\cdots1]'$) 2. $Y'M^0Y=\hat Y'M^0\hat Y+\hat\varepsilon'\hat\varepsilon\Rightarrow SST=SSE+SSR$

$R^2=1-\frac{SSR}{SST}=1-\frac{\sum\hat\varepsilon_i^2}{\sum(Y_i-\bar Y)^2}$, $\bar R^2=1-\frac{SSR/(n-k)}{SST/(n-1)}=1-\frac{n-1}{n-k}(1-R^2)$

Finite Sample Properties: 1. $E(\hat\beta|X)=\beta$ 2. $Var(\hat\beta|X)=\sigma^2(X'X)^{-1}$

(without asy.4) $Var(\hat\beta|X)=(X'X)^{-1}X'\Sigma X(X'X)^{-1}$, where $Var(\varepsilon|X)=\Sigma$

Proof for BLUE: let $\tilde\beta=AY$ be a linear estimator of β, where A is a k×n matrix

unbiasedness: $E(\tilde\beta|X)=E[A(X\beta+\varepsilon)|X]=E[AX\beta+A\varepsilon|X]=AX\beta+AE[\varepsilon|X]=\beta$

this follows iff $AX=I_k$

Variance: $Var(\tilde\beta|X)=Var(AX\beta+A\varepsilon|X)=Var(A\varepsilon|X)=AVar(\varepsilon|X)A'=\sigma^2AA'$

$Var(\tilde\beta|X)-Var(\hat\beta|X)=\sigma^2AA'-\sigma^2(X'X)^{-1}=\sigma^2[AA'-(X'X)^{-1}]=\sigma^2[AA'-AX(X'X)^{-1}X'A']$

$=\sigma^2A[I_n-X(X'X)^{-1}X']A'=\sigma^2AMA'=\sigma^2AM(AM)'$

$Cov(\hat\beta,\hat\varepsilon|X)=E((\hat\beta-\beta)(M\varepsilon)'|X)=E((X'X)^{-1}X'\varepsilon\varepsilon'M|X)=(X'X)^{-1}X'E(\varepsilon\varepsilon'|X)M=\sigma^2(X'X)^{-1}X'M=0$

$E(s^2|X)=\frac{1}{n-k}E(\varepsilon'M\varepsilon|X)=\frac{1}{n-k}tr\,E(\varepsilon'M\varepsilon|X)=\frac{1}{n-k}E[tr(M\varepsilon\varepsilon')|X]=\frac{1}{n-k}tr[ME(\varepsilon\varepsilon'|X)]=\frac{1}{n-k}tr(M\sigma^2I_n)=\frac{\sigma^2}{n-k}tr(M)=\sigma^2$

- consistency: $\hat\beta-\beta=(X'X)^{-1}X'u=(\frac1n\sum X_iX_i')^{-1}(\frac1n\sum X_iu_i)=0$ ($u\sim(0,\sigma^2\Sigma)$)
- normality: $\sqrt n(\hat\beta-\beta)\xrightarrow{d}N(0,Q^{-1}VQ^{-1})$, where $Q=E(X_iX_i')$, $V=E(X_iX_i'u_i^2)$

  $\hat\beta=\beta+(X'X)^{-1}X'u$, $\sqrt n(\hat\beta-\beta)=(\frac1nX'X)^{-1}(\frac{1}{\sqrt n}X'u)$ or $\hat\beta|X=(X'X)^{-1}(X'u)$

  $\frac1nX'X=\frac1n\sum X_iX_i'\xrightarrow{p}Q$ (WLLN)

  $\frac{1}{\sqrt n}X'u=\frac{1}{\sqrt n}\sum X_iu_i\xrightarrow{d}N(0,V)$ (CLT) $=E(X_iX_i'u_i^2)$

  $\Rightarrow\sqrt n(\hat\beta-\beta)\xrightarrow{d}Q^{-1}N(0,V)=N(0,Q^{-1}VQ^{-1})$ (Slutsky Theorem)