CandyOre 22 秋 考试 A4 纸

Confidence Interval

90%: 1.65 95%: 1.96 99%: 2.58

Chapter 9 Threat to Internal Validity

a. Omitted Variable bias

(1) Condition

1. determinant of Y
2. correlated with at least one included regressor

$$\beta_2 \cdot \frac{Cov(X_1, X_2)}{Var(X_2)}$$

(2) Solution

1. include 2. control 3. panel 4. IV
5. Randomized Controlled Exp. $E(u|X=x)=0$

b. Functional Form Mipecification

c. Errors-in-Variables bias

(1) Classical measurement error

1. $\tilde{X}_i = X_i + v_i$ where v_i is mean-zero random noise with $corr(X_i, v_i) = 0$ and $corr(u_i, v_i) = 0$.
2. biased towards zero: $\beta_1 \to \left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_v^2}\right)\beta_1$.

(2) Best guess

1. $\tilde{X}_i = E(X_i|W_i)$, where $E(u_i|W_i)=0$.
2. $cov(\tilde{X}_i,\ \tilde{u}_i)=0$, so $\hat{\beta}_1$ is unbiased.

(3) Solution

1. better data 2. specific model 3. IV

d. Missing Data and Sample Selection bias

(1) Missing at random: no bias.
(2) Missing based on one of the X, no bias.
(3) Missing based on Y or u: Sample selection bias.

1. influence the availability of data
2. is related to Y. Now $Corr(X_i, u_i) \neq 0$
3. Solution: i. collect .. avoid .. ii. Rand.Cont.Exp. iii. model

e. Simultaneous Causality bias

(1) $Corr(X_i, u_i) \neq 0$ biased & inconsistent.
(2) Solution: i. Rand.CE ii. Model iii. IV

Chapter 10 Panel Data

$$(X_{1,it}, X_{2,it}, \dots, X_{k,it}, Y_{it}),\ i = 1, \dots, n,\ t = 1, \dots, T$$

a. Fixed effect across states

(1) $n-1$ binary regressor form

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \ \dots\ + \gamma_n D_{ni} + u_{it}$$

(2) Fixed effects form (entity-demeaned)

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

i.e. $\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$ where $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$

and $\tilde{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$

(3) Changes specification $T=2$

$$Y_{it1} - Y_{it2} = \beta_1(X_{it1} - X_{it2}) + (1-1)\beta_2 Z_i + u_i'$$

(4) OV bias from variables that vary across states.

b. Fixed effect across time

(1) $T-1$ binary regressor

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \ \dots \delta_T BT_t + u_{it}$$

(2) Time effect version (year-demeaned)

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}$$

(3) OV bias from variables that vary over time.

c. Clustered standard errors

(1) u_{it} is serially correlated.
(2) $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $(\overline{Y_1}, \dots \overline{Y_n})$ i.i.d.

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} \bar{Y}_i \to^d N(0,\ \sigma_{\bar{Y}_i}^2/n)$, where $\sigma_{\bar{Y}_i}^2 = \text{var}(\bar{Y}_i)$.

(3) Clustered SE $\bar{Y} = \sqrt{\frac{s_{\bar{Y}_i}^2}{n}}, s_{\bar{Y}_i}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \bar{Y})^2$

Chapter 11 Binary Dependent Variable

a. Models

(1) Linear Probability Model

$$\Pr(Y=1|X) = \beta_0 + \beta_1 X$$

(2) Probit Regression Model

$$\Pr(Y=1|X) = \Phi(\beta_0 + \beta_1 X)$$

Φ is the cumulative normal distribution function.

$z = \beta_0 + \beta_1 X$ is the "z-value" of the probit model.

(3) Logit Regression Model

$$\Pr(Y=1|X) = F(\beta_0 + \beta_1 X)$$

$$F(\beta_0 + \beta_1 X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$$

b. Maximum Likelihood Estimator

$$\Pr(Y_1 = y_1|X_1) = \Phi(\beta_0 + \beta_1 X_1)^{y_1}[1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-y_1}$$

大 sample: Consistent, normally distributed, efficient.

c. Measures of Fit

(1) Fraction correctly predicted

$$(>50\%)\&(Y_i = 1) + (<50\%)\&(Y_i = 0)$$

(2) Pseudo-R: improvement in the value of the log likelihood, relative to having no X 's.

Chapter 12 Instrumental Variables

a. Valid IV and Usage

(1) Validity

Relevance: $corr(Z_i, X_i) \neq 0$

Exogeneity: $corr(Z_i, u_i) = 0$

(2) Two Stage Least Square (TSLS)

1. Reg X on Z: $X_i = \pi_0 + \pi_1 Z_i + v_i$
2. Reg Y on $\hat{X}$: $Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$

(3) Algebraic

1. $cov(Y_i, Z_i) = \beta_1 cov(X_i, Z_i)$
2. IV estimator replaces pop cov with sample cov

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}} \to^p \beta_1$$

(4) Reduced Form

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$
$$Y_i = \gamma_0 + \gamma_1 Z_i + w_i$$
$$\beta_1 = \gamma_1/\pi_1$$

(5) Inference

$$\hat{\beta}_1^{TSLS} \sim N\left(\beta_1,\ \sigma_{\hat{\beta}_1^{TSLS}}^2\right), \sigma_{\hat{\beta}_1^{TSLS}}^2 = \frac{1}{n}\frac{var[(Z_i - \mu_Z)u_i]}{[cov(Z_i, X_i)]^2}$$

b. General IV

(1) Setting

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \dots + \beta_{k+r} W_{ri} + u_i$$

- Exogenous variables $E(u_i|W_i, Z_i) = E(u_i|W_i)$
- IV $(Z_1, \dots, Z_m)$

(2) Identification

- Overidentified $m > k$
- Underidentified $m < k$

(3) Validity

- $corr(Z_{1i}, u_i) = 0, \dots, corr(Z_{mi}, u_i) = 0$
- no perfect multicollinearity in second stage

c. Weak IV

(1) Condition: First-stage F-statistic < 10.

(2) Anderson-Rubin confidence interval

- $\tilde{u}_i = Y_i - \beta_{1,0} X_i$
- Regress $\tilde{u}_i$ on $W_{1i}, \dots, W_{ri}, Z_{1i}, \dots, Z_{mi}$
- AR test is heterosk-robust F-statistic on $Z_{?i} = 0$
- 95% interval: all β_1 not rejected at 5% AR test

d. J-test for overidentifying exogeneity

(1) Above AR test but $\tilde{u}_i = Y_i - \hat{Y}_i$
(2) $J = mF \sim \chi^2(m-k)$
(3) Reject null that all IV are exogeneity.

Chapter 13 Experiments & Quasi-Exp

a. Experiments

$$\begin{aligned} Y_i &= Y_i(1)X_i + Y_i(0)(1-X_i) \\ &= E[Y_i(0)] + [Y_i(1) - Y_i(0)]X_i + [Y_i(0) - E(Y_i(0))] \\ &= \beta_0 + \beta_{1i} X_i + u_i \end{aligned}$$

(1) X_i is randomly assigned, $E(u_i|X_i) = 0$.
(2) β_1 is **difference estimator**.
(3) Threats to Internal Valitdiy

1. Failure to randomize 2. Partial compliance
3. Attrition (subjects drop out) 4. Exp bias

(4) Threats to External Validity

1. Nonrepresentative sample / treatment
2. Depend on scale (general equilibrum effect)

b. Quasi-Experiments

(1) Type I. X is as if randomly assigned

- Differences-in-differences estimator β_1

$$\Delta Y_i = \beta_0 + \beta_1 X_i + u_i$$

(2) Type II. Z which influences X is as if randomly ass

- IV

(3) Regression discontinuity design

- Shape: above threshold w_0 gets treatment
- Fuzzy: w_0 influences probability of treatment

(4) Threats to Internal Valitdiy

1~4. = Exp. threats 5. IV invalidity

c. Heterogeneous Population

(1) OLS estimates Average Treatment Effect (ATE)

$$\beta_1 = E(\beta_{1i}) = E[Y_i(1) - Y_i(0)]$$

(2) IV estimates Local ATE (LATE)

$$Y_i = \beta_0 + \beta_{1i} X_i + u_i$$
$$X_i = \pi_0 + \pi_{1i} Z_i + v_i$$

$$\hat{\beta}_1^{TSLS} \xrightarrow{p} \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})} = \text{LATE}$$

$$\text{LATE} = \text{ATE} + \frac{\text{cov}(\beta_{1i}, \pi_{1i})}{E(\pi_{1i})}$$

1. LATE is effect for those most influenced by Z
2. IV extimates $E(\beta_{1i})$ if

- If $\beta_{1i} = \beta_1$ (no heterogeneity in equation of interest)
- If $\pi_{1i} = \pi_1$ (no heterogeneity in first stage equation)
- If β_{1i} and π_{1i} vary but are independently distributed.

Chapter 14 Many Regressor & Big Data

(1) Many regressors, OLS will produce poor out-of-sample(OOS) prediction.
(2) Bias to zero leads to smaller MSPredictionError.
(3) Ridge Reg: $S^{Ridge}(b; \lambda_{Ridge}) = \sum_{i=1}^{n}(Y_i - b_1 X_{1i} - \dots - b_k X_{ki})^2 + \lambda_{Ridge}\sum_{j=1}^{k} b_j^2$
(4) Lasso Reg: $S^{Lasso}(b; \lambda_{Lasso}) = \sum_{i=1}^{n}(Y_i - b_1 X_{1i} - \dots - b_k X_{ki})^2 + \lambda_{Lasso}\sum_{j=1}^{k}|b_j|$
(5) Principle Component with p component $PC_j = \text{maxvar}\left(\sum_{i=1}^{k} a_{ji} X_i\right)$, subject to $\sum_{i=1}^{k} a_{ji}^2 = 1$
(6) Parameter etimated by minimizing m-fold cross-validated estimate of the MSPE.

Chapter 15 Time Series Forecasting

a. Notation

(1) Data set: $\{Y_1, \dots, Y_T\}$
(2) Lag: Y_t's j^{th} lag is Y_{t-j}
(3) Difference: $\Delta Y_t = Y_t - Y_{t-1}$
(4) Difference of the logarithm: $\Delta\ln(Y_t) = \ln(Y_t/Y_{t-1})$
(5) j^{th} autocovariance $cov(Y_t,\ Y_{t-j})$
(6) j^{th} autocorrelation = serial correlation coefficient $corr(Y_t, Y_{t-j}) = \frac{cov(Y_t, Y_{t-j})}{\sqrt{var(Y_t)\,var(Y_{t-j})}}$.
(7) j^{th} sample autocorrelation $\frac{\widehat{\text{cov}(Y_t, Y_{t-j})}}{\widehat{\text{var}(Y_t)}}$
(8) Stationary: $(Y_{s+1}, \dots, Y_{s+T})$ doesn't depend on s.
(9) Forecast: $Y_{T+h|T}$. Multi-step ahead: $h > 2$.
(10) MSForecastError $E\left[\left(Y_{T+1} - \hat{Y}_{T+1|T}\right)^2\right]$
(11) pth order autoregression AR(p):

- Reg Y_t on $Y_{t-1}, \cdots, Y_{t-p}$
- $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$

(12) Autoregressive Distributed Lag ADL(p,r)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \dots + \delta_r X_{t-r} + u_t$$

b. Estimate Forecast Error

(1) Approximate by Standard Error of Regerssion

$$\widehat{MSFE}_{SER} = s_{\hat{u}}^2 = \frac{SSR}{T-p-1}.$$

(2) Final Prediction Error

$$\widehat{MSFE}_{FPE} = \left(\frac{T+p+1}{T}\right)s_{\hat{u}}^2 = \left(\frac{T+p+1}{T-p-1}\right)\left(\frac{SSR}{T}\right).$$

(3) Pseudo out-of-sample (POOS)

$$\widehat{MSFE}_{POOS} = \frac{1}{P}\sum_{s=T-P+1}^{T}\tilde{u}_s^2, \tilde{u}_{s+1} = Y_{s+1} - \hat{Y}_{s+1|s}.$$

(4) $\text{RMSFE} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$.
(5) 95% forecast interval $\hat{Y}_{T+1|T} \pm\ 1.96 \times \widehat{RMSFE}$

c. Information Criteria(IC)

(1) Choose p in AR(p)
(2) Bayes IC $\text{BIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$
(3) Akaike IC $\text{AIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$

- AIC isn't consistent, overestimate p.

(4) BIC for ADL(p,r)

$\text{BIC}(K) = \ln\left(\frac{SSR(K)}{T}\right) + K\frac{\ln T}{T}, K = p + r + 1.$

d. Nonstationarity I: Trends

(1) Deterministic trend is nonrandom function of time
(2) Stochastic trend is random and varies over time

1. Random walk $Y_t = Y_{t-1} + u_t$ where u_t not autocorr.
   - $Y_{T+h|T} = Y_T, var(Y_t) = t\sigma_u^2$.
2. Random walk with drift $Y_t = \beta_0 + Y_{t-1} + u_t$
   - $Y_{T+h|T} = Y_T + \beta_0 h$.
3. A unit root in AR(p) = AR(p-1) in first diff
   $\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_{p-1}\Delta Y_{t-p+1} + u_t$
   - unit root for $1 - \beta_1 z_1 - \dots - \beta_p z_p = 0$
4. Should use AR(p-1) in first diff version, or
   - AR coefficient strongly biased towards 0
   - t-statistics don't have a normal distribution
   - Y and X with random walk cause fake corr.

(3) Stochastic trend detection

1. Plot: persistent long-run movements
2. Dickey-Fuller test for unit root

- $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$
- $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$
- H_0: $\delta = 0$ (that is, $\beta_1 = 1$) v. H_1: $\delta < 0$
- Compute the t-statistic testing $\delta = 0$
- Reject if the DF t-statistic < critical value
- Can include t as a regressor in ΔY_t

e. Nonstationarity II: Breaks

(1) Break date is known

1. include interacted term
2. perform Chow test on all the interacted term
3. Check F-statistic

(2) Unknown: Quandt Likelihood Ratio(QLR) test

1. set break date τ
2. QLR statistic is the maximum of all Chow F
3. QLR distribution $\max_{a \le s \le 1-a}\left(\frac{1}{q}\sum_{i=1}^{q}\frac{B_i(s)^2}{s(1-s)}\right)$
4. Conventional inner 70% of the sample.

(3) POOS for end-of-sample forecast breakdown

Chapter 16 Dynamic Causal Effect

2021-2022

(1) Interpretion
- holding xxx constant.
- IV for LATE.

(2) Selection Bias

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

(3) Heterogeneous J-test

Recall that in a heterogeneous world, different instruments estimate different LATEs. Thus, just because the J-stat reject the null hypothesis doesn't necessarily mean that either is invalid. Rather, the different instruments might just be picking up the fact that different instruments have different LATEs.

(4) Control group in DID

Here the implicit control group is countries that did not experience a transition from a non- democratic to a democratic government in 1985. This control group might not provide a good counterfactual for what would have happened to countries that experienced a democratic transition had they not experienced such a transition.

(5) Source of variation.
- cross-sectional (between country) variation
- within-country variation (different time)

(6) Time series confidence interval: use RMSFE.

(7) Newey-West rule of thumb

$m=0.75T^{\frac{1}{3}}$ round up, T is observation size.

2019-2020

(1)

$E[Trust_i|Int_i=1, Tol_i=1]$-$E[Trust_i|Int_i=0, Tol_i=1]$ by random assignment of Int_i this is equal to $E[Trust_i(1)|Tol_i(1)=1]$-$E[Trust_i(0)|Tol_i(0)=1]$ or $E[Trust_i(1)-Trust_i(0)|Tol_i(1)=1]$+ $E[Trust_i(0)|Tol_i(1)=1]$- $E[Trust_i(0)|Tol_i(0)=1]$. Or a causal effect plus selection bias

(2)

5) (4 points)Another observer wants to allow for the cubic relationship between test scores and trust to be different for individual with scores below 8.7 and those who got an 8.7 or higher. How would you modify regression (2) to allow for such a relationship? Be specific.

Here the modified regression would be $Trust_i=\beta_0+\beta_1 D_i+\beta_2(Score_i-8.7)+\beta_3(Score_i-8.7)^2+\beta_4(Score_i-8.7)^3+\beta_5(Score_i-8.7)*D_i+\beta_6(Score_i-8.7)^2*D_i+\beta_7(Score_i-8.7)^3*D_i+u_i$,

Subtracting the 8.7 is to ensure that the coefficient β_1 gives the jump at the value 8.7 which is the RDD coefficient. If you do not make this normalization you would get the jump at 0 which would give you the incorrect answer. Correctly "re-centering" the regression was worth two point.

(3)

6) (4 points)Using the regression output in equation (6), discuss the internal validity of this causal effect. Be specific.

For the results to be internally valid, the change in sales for fast food chains that only operated in US areas outside of New England must provide a valid counterfactual for what would have happened to sales for chains that operated in New England had they not had their food contaminated (2 points). In particular, both groups would need to follow a common trend absent treatment (1 point). Although this assumption is impossible to test, the coefficient on $NewEngland_i$ shows that fast food chains that operated in New England were larger than the control group in 1995 (although since we are not given the standard error it is impossible to know the extent to which this is just noise). Thus, it may have been that counterfactual change would have also been different for these larger firms (1 point)

(4) ADL(p,r)
- use BIC to choose p
- use Granger causality test (F-test) to choose r

(5) m-fold cross-validation, for the case m = 10:
1. Estimate the model on 90% of the data and use it to predict the remaining 10%
2. Repeat this on the remaining 9 possible subsamples (so there is no overlap on the test samples).
3. Estimate the MSPE using the full set of out-of-sample predictions

2020-2021

(1) Conditional mean independence.

$$E(u|X,W)=E(u|W)$$

means given W in the regression, X's estimator is unbiased.

(2) SE(ax)=aSE(x)

(3) The difference in these probabilities is zero if and only if β1 = 0 in the probit regression.

(4) BIC selects the true model with probability 1 in large samples.

(5)

3) (3 points) The Newey-West standard errors for model (5) are quite similar in magnitude to the heteroskedasticity-robust standard errors for model (4). Why?

The NW standard errors adjust for serial correlation in the error term. If the errors are serially uncorrelated, then those adjustments are not needed and the NW and HR SEs will be close to each other. This is evidently the case in regression (4), that is, the errors in regression (4) are approximately serially uncorrelated. This is, in fact, what is expected in a sufficiently long autoregression: because lags of *y* are included, the forecast error should itself be unforecastable using lagged *y* (if it were predictable, the autoregression would predict it – but then that predictable component would be in the predicted value, not in the error term.) So in fact NW SEs are not needed in an autoregression.