Suparat Patarasupanit (FT27)
6684069326

Financial Econometrics : Problem set 2

I

1. Consider each of the following statements is true, is false, or has not sufficient information to make a conclusion.
   a. Let $Z1$ and $Z2$ be valid instruments. A linear combination of the two instruments are also a valid instrument.
   b. Measurement error in a regressor can cause endogenous problem for the OLS estimator.
   c. MLE is a consistent estimator for linear regression model.
   d. If the error term is not mean-independent of a regressor, then the OLS estimator is not consistent.
   e. OLS estimator is inconsistent when serial correlation is present.
   f. Newey-West Standard Error can be used when data contain heteroskedasticity.
   g. ARMA process is weak stationary
   h. Random Walk process is weak stationary
   i. Under the Random Effect assumption, OLS is consistent.
   j. OLS estimator is inconsistent in with the presence of ARCH.

(a) There is no sufficient information to conclude that a linear combination of z_1 and z_2 is also a valid instrument. For a linear combination $\alpha_0 + \alpha_1 z_1 + \alpha_2 z_2$, we need to check first that $\alpha_1 \text{cov}(z_1, x) + \alpha_2 \text{cov}(z_2, x) \neq 0$. In other words, linear combination of z_1 and z_2 has to be correlated with endogenous variable (x) as well.

(b) The statement is true. Measurement error occurs because we cannot observe the regressor (x). We can only observe and use $\tilde{x}$ as regressor which also has error in itself ($\tilde{x} = x + \overset{\text{error}}{\epsilon}$). Thus, endogenous problem occurs because $\text{cov}(\tilde{x}, \epsilon) \neq 0$.

(c) There is no sufficient information to conclude that MLE is a consistent estimator for linear regression model unless the normality assumption is true $\{ u_i | x_i \sim iid\ N(0, \sigma^2) \}$.

(d) There is no sufficient information to conclude the statement. OLS estimator is consistent if $\text{cov}(x, u) = 0$ but it does not require error term to be mean independent.

(e) False, serial correlation occurs when $\text{cov}(u_t, u_s) \neq 0$ for some period $s \neq t$ but it does not imply that $\text{cov}(x, u) \neq 0$. Thus, we cannot conclude that OLS estimator is inconsistent.

(f) True, Newey-West Standard Error can be used when data contain heteroskedasticity.

(g) False, ARMA can be non-stationary. To illustrate, random walk process or ARMA(1,0) with $\rho = 1$ is non-stationary.

(h) False, random walk process is non-stationary.

(i) True, under random effect assumption, OLS is consistent as $\text{cov}(x_{it}, \alpha_i) = 0$.

(j) There is no sufficient information to conclude the statement because ARCH is a heteroscedasticity model and it does not determine whether $\text{cov}(u, x) = 0$ or not.

II

2. In which of the following situations, can OLS be a consistent estimator? Explain.
   a. Fixed Effect
   b. Random Effect
   c. Omitted variable
   d. Simultaneous equations
   e. Measurement error

topic 10

ⓐ OLS cannot be a consistent estimator for Fixed Effect because Fixed Effect has endogeneity problem as $cov(X, \alpha) \neq 0$. (α → unobserved heterogeneity & left in error term)

ⓑ OLS is consistent estimator for Random Effect because $cov(X, \alpha) = 0$. However, there is serial correlation problem.

topic 6 pg 5

ⓒ OLS is consistent estimator when omitted variable bias occur if and only if $E[XW] = 0$ (W → omitted variable) or the omitted variable is uncorrelated with each of the other regressors.

topic 6 pg. 11

ⓓ OLS cannot be a consistent estimator for simultaneous equations. For example, we can solve for X from simultaneous equations including $\beta_0 + \beta_1 X + U = Y = \gamma_0 + \gamma_1 X + \eta$ and get $cov(X, \eta) \neq 0$ and $cov(X, U) \neq 0$.

topic 6

ⓔ OLS can be a consistent estimator for the measurement error if $cov(\tilde{X}, \varepsilon) = 0$ or no endogeneity problem for the model $y = \beta_0 + \beta_1 \tilde{X} + (U - \beta_1 \varepsilon)$.

III

3. Let X be a discrete random variable with the following p.m.f.

* λ is unknown parameter of interest

$$P\{X = x\} = \frac{\lambda^x \exp(-\lambda)}{x!}$$

Then we say that X is distributed with Poisson Distribution with parameter λ. This is a well-known distribution proposed by a French Mathematician, Poisson (1837), and is used to express the probability of the number of events occurring in a given fixed interval of time, provided that each event happens independently of the others. Note that X can take values 0, 1, 2, ..., and all positive integers. For example, X may be the number that you receive a phone call in a day, given that the chance that one person calls you doesn't affect the chance that the others call you.

a. Suppose that you know that your data $X_1, \dots, X_n \sim iid\ Poisson(\lambda)$ but you don't know λ and want to estimate it by MLE. Write down the likelihood function and log-likelihood function of observing the data set.

b. Prove that $\hat{\lambda}_{MLE} = \bar{X}_n$.

ⓐ $X_1, \dots, X_n \sim$ iid which has the p.m.f. $P(X;\lambda)$

$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \times \dots \times P\{X_n = x_n\}$

▸ Likelihood function:

$$L(X;\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{X_i} \exp(-\lambda)}{X_i!} \right)$$

▸ Log - likelihood function

$$\log(X;\lambda) = \log\left(\prod_{i=1}^{n} \left(\frac{\lambda^{X_i} \exp(-\lambda)}{X_i!} \right) \right)$$

note: $\log(XY) = \log X + \log Y$

$$= \sum_{i=1}^{n} \log\left(\frac{\lambda^{X_i} \exp(-\lambda)}{X_i!} \right)$$

note: $\log\left(\frac{X}{Y}\right) = \log X - \log Y$

$$= \sum_{i=1}^{n} \left(\log(\lambda^{X_i} \exp(-\lambda)) - \log(X_i!) \right)$$

$$= \sum_{i=1}^{n} \left(\log(\lambda^{X_i}) + \log(\exp(-\lambda)) - \log(X_i!) \right)$$

$$= \sum_{i=1}^{n} \left(X_i \log\lambda + (-\lambda) - \log(X_i!) \right)$$

ⓑ Prove that $\hat{\lambda}_{MLE} = \bar{X}_n$

To find MLE, we take derivative on log-likelihood function (first-order condition (FOC))

$$\hat{\lambda}_{MLE} = \frac{\partial}{\partial\lambda} \log(X;\lambda) = \sum_{i=1}^{n} \left(\frac{X_i}{\lambda} - 1 - \cancel{\frac{1}{X_i!}}^{\,0} \right) = \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n = 0$$

$$\frac{1}{\lambda} \sum_{i=1}^{n} X_i = n$$

$$\therefore \lambda = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n$$

topic 8

$\boxed{IV}$

4. Suppose that we want to study the effect of policy interest rate, denoted by int (X), on inflation rate as denoted by inf (Y). Suppose that we use the finite distributed lag model and get the following result.

Dependent Variable inf			lag 1 $t-1$	lag 2 $t-2$
Regressors	_con	int	$l.int$	$l2.int$
Coefficient	1.633	0.479	-0.157	-0.323
Std. Error	0.65	0.22	0.004	0.023

a. What is the immediate impact propensity of the policy rate on inflation?
b. Suppose that the change in interest rate is temporary, what are the effect of policy rate change on inflation rate after one, two, and three periods?
c. What is the long-run multiplier of the policy rate on inflation?
d. Suppose that the central bank set interest rate by also looking at contemporaneous inflation rate. Do you think we can use OLS to estimate this model? If not, what should be a proper estimator to use?

pg. 36 The model : $inf_t = \beta_0 + \beta_1 int_t + \beta_2 int_{t-1} + \beta_3 int_{t-2} + u_t$

ⓐ The immediate impact propensity is $inf_t - inf_{t-1}$.

$$inf_t = \beta_0 + \beta_1 int_t + \beta_2 int_{t-1} + \beta_3 int_{t-2} = \beta_0 + \beta_1$$
$$inf_{t-1} = \beta_0 + \beta_1 int_{t-1} + \beta_2 int_{t-2} + \beta_3 int_{t-3} = \beta_0$$

∴ The immediate impact propensity $= inf_t - int_{t-1} = \beta_0 + \beta_1 - \beta_0 = \beta_1 = 0.479$

ⓑ Temporary change : government increases policy interest rate 1 unit from c to c+1 only in period t and revert back.

Predicted inflation :

$$inf_{t-1} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 c$$
$$inf_t = \beta_0 + \beta_1 (c+1) + \beta_2 c + \beta_3 c$$
$$inf_{t+1} = \beta_0 + \beta_1 c + \beta_2 (c+1) + \beta_3 c$$
$$inf_{t+2} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 (c+1)$$
$$inf_{t+3} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 c$$

t	int_t	int_{t-1}	int_{t-2}	$\Rightarrow inf_t$
-2	0	0	0	
-1	0	0	0	β_0
0	1	0	0	$\beta_0 + \beta_1$
1	0	1	0	$\beta_0 + \beta_2$
2	0	0	1	$\beta_0 + \beta_3$
3	0	0	0	β_0

∴ Effect of policy interest rate change on inflation rate

- After 1 period $(t+1)$ is $\beta_0 + \beta_2 - \beta_0 = \beta_2 = -0.157$
- After 2 periods $(t+2)$ is $\beta_0 + \beta_3 - \beta_0 = \beta_3 = -0.323$
- After 3 periods $(t+3)$ is $\beta_0 - \beta_0 = 0$ or no effect

Topic 8

IV

c. What is the long-run multiplier of the policy rate on inflation?

d. Suppose that the central bank set interest rate by also looking at contemporaneous inflation rate. Do you think we can use OLS to estimate this model? If not, what should be a proper estimator to use?

pg. 37

ⓒ Long-run multiplier is $\beta_1 + \beta_2 + \beta_3 = 0.479 - 0.157 - 0.323 = -0.0001$

∴ If we increase policy rate by 1 unit and keep it permanently, inflation rate in the long run we decrease by 0.001 unit.

Topic 6 ⓓ Central bank sets interest rate by also looking at contemporaneous inflation. This means that, central bank wants to study the effect of inf_t on int_t (causation). This can cause simultaneous equations problem and OLS will not be a consistent estimator for this model. We can solve this problem by using 2SLS estimator instead.

V

5. Suppose that we have a sample of working-age Thai people with college degrees, and we want to study factors contributing to the probability that they invest in the Stock Exchange of Thailand. Let SET be a dummy variable taking value 1 if a person invests in SET; inc is the person's annual income in units of thousand Baht; $exper$ is the person's working experience in years; age is the person's age in years; and fin is a dummy variable taking 1 if the person graduated with a finance-related degree. Suppose that we use the Logit model to estimate and get the following result:

* Regress SET on all independent X

Dependent variable: SET					
Regressors	Constant	inc	$exper$	age	fin
$\hat{\beta}$	-2.74	0.0112	0.0079	-0.0014	1.280
$se(\hat{\beta})$	(0.12)	(0.001)	(0.003)	(0.028)	(0.56)

a. What is the probability predicted by the estimated model that a 30 years old person with a Finance degree and 7 years of work experience who earn 10,000 Baht a month would invest in SET?

$\Downarrow$

$\frac{10{,}000 \times 12}{1000} = 120$ Thousand Baht a year

* Use LOGIT model

ⓐ $Pr\{SET = 1 \mid x\} = \dfrac{e^{\hat{\beta}_0 + \hat{\beta}_1 inc + \hat{\beta}_2 exper + \hat{\beta}_3 age + \hat{\beta}_4 fin}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 inc + \hat{\beta}_2 exper + \hat{\beta}_3 age + \hat{\beta}_4 fin}}$

for the given person

$$= \frac{e^{-2.74 + 0.0112(120) + 0.0079(7) - 0.0014(30) + 1.280(1)}}{1 + e^{-0.1027}}$$

-0.1027

$$= \frac{0.9024}{1 + 0.9024}$$

$$= 47.43\%$$

$* \; x\beta = \beta_0 + \beta_1 inc + \beta_2 exper + \beta_3 age + \beta_4 fin$

$\boxed{V}$

$\frac{120{,}000}{1000} = 120$ thousand Baht a year

b. For a person ages 25 years old with 2 years of work experience earning 120,000 Baht a year, what is the marginal effect of a finance degree on the probability that he invests in SET?

(age: 25; exper: 2)

c. For a person ages 25 years old with a finance degree and 2 years of work experience earning 120,000 Baht a year, what is the effect of income level on the probability that he invests in SET?

d. Now suppose that the given result is from Probit instead of Logit. What is the probability predicted by the estimated model that a 30 years old person with a Finance degree and 7 years of work experience who earns 10,000 Baht a month would invest in SET?

ⓑ • If he does not have finance degree (fin = 0)

$$Pr\{SET=1 \mid X\} = \frac{e^{-2.74+0.0112(120)+0.0079(2)-0.0014(25)+1.280(0)}}{1+e^{-1.4152}}$$

-1.4152

$= 19.54\%$

• If he has finance degree (fin = 1)

$$Pr\{SET=1 \mid X\} = \frac{e^{-1.4152+1.280}}{1+e^{-0.1352}} = 46.63\%$$

-0.1352

$46.63\% - 19.54\%$

∴ Marginal effect of a finance degree on the proability that he invests in SET = 27.09%

ⓒ Effect of income level on $Pr\{SET \mid X\}$ is

$$\frac{\partial}{\partial inc}\left(Pr\{SET=1 \mid X\}\right) = \frac{\partial}{\partial inc}\left\{\frac{e^{x\beta}}{1+e^{x\beta}}\right\}$$

$$\boxed{\frac{\partial}{\partial inc}(1+e^{x\beta}) = \frac{\partial}{\partial inc} e^{x\beta}}$$

$$= \frac{(1+e^{x\beta})\frac{\partial}{\partial inc}(e^{x\beta}) - (e^{x\beta})\frac{\partial}{\partial inc}(1+e^{x\beta})}{(1+e^{x\beta})^2}$$

$$= \frac{\frac{\partial}{\partial inc}(e^{x\beta})\,(\cancel{1+e^{x\beta}-e^{x\beta}})^{\to 1}}{(1+e^{x\beta})^2}$$

$$= \frac{e^{x\beta}\cdot\frac{\partial}{\partial inc}(x\beta)}{(1+e^{x\beta})^2}$$

$$= \left\{\frac{e^{x\beta}}{(1+e^{x\beta})^2}\right\} \times \beta_1$$

<(Cont.)>

ⓒ For age = 25, fin = 1, exper = 2, inc = 120 thousand Baht a year

↳ From ⓑ, $x\hat{\beta} = -0.1352$

* Effect of income level on $Pr\{SET = 1 | x\}$ for the given person is:

$$\frac{\partial}{\partial inc}\left(Pr\{SET = 1 | x\}\right) = \left\{\frac{e^{x\hat{\beta}}}{(1 + e^{x\hat{\beta}})^2}\right\} \times \hat{\beta}_1$$

$$= \left\{\frac{e^{-0.1352}}{(1 + e^{-0.1352})^2}\right\} \times 0.0112$$

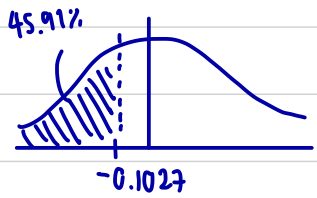
$$= 0.2489 \times 0.0112$$

$$= 0.002787 \quad \text{or} \quad 0.2787\%$$

ⓓ age = 30, fin = 1, exper = 7, inc = $\frac{10{,}000 \times 12}{1000}$ = 120 thousand Baht per year

PROBIT model:

$$Pr\{SET = 1 | x\} = \Phi(\underbrace{\beta_0 + \beta_1 inc + \beta_2 exper + \beta_3 age + \beta_4 fin}_{\text{same as ⓐ} = -0.1027})$$



$$= \Phi(\underbrace{-0.1027}_{z})$$

obtain this from z calculator

$$= 0.4591 \quad \text{or} \quad 45.91\%$$

VI

6. Suppose a researcher has a sample of 200 monthly observations on the stochastic process $\{Y_t\}$ and wants to fit the data with an ARMA model. In order to select the model, he computes the ACF and PACF as follows:

k	1	2	3	4	5	6	7	8	9
ACF	0.83	0.71	0.60	0.45	0.44	0.35	0.29	0.20	0.11
PACF	0.83	0.16	-0.09	0.05	0.04	-0.05	0.01	0.10	-0.03

a The researcher decides not to use the MA model to fit the data. Based on the given correlogram, why do you think he makes such decision?

ⓐ In my opinion, he did not use MA model because we do not observe MA process from ACF result. In other words, to use the MA(q) model, ACF should decrease and then equal to zero from $k = q$ lag onwards. However, ACF is gradually decreasing and approaching 0 but not equal to 0 (AR(p) process).

b Suppose that the researcher suspects that AR(2) may be better. So, he runs AR(2) and get the following result:

$$Y_t = 0.74Y_{t-1} + \hat{\rho}_2 Y_{t-2} + e_t$$

What is the numerical value of $\hat{\rho}_2$ in the above equation?

ⓑ We use PACF for AR(p).

AR(2) → k = 2
↑
lags

∴ $\hat{\rho}_2 = 0.16$

c Conduct a hypothesis testing to see if he would prefer AR(1) or AR(2). State the null and alternative hypotheses, test statistics, distribution of the test statistics, critical value at 5% significance level, p-value, and the conclusion of the test whether AR(1) or AR(2) is preferred.

Topic 8
pg 17

ⓒ • Hypothesis testing for AR(1) or $\rho = 1$?

AR(1) : $y_t = \rho_{11} y_{t-1} + e_t$

two-tailed test → • Hypothesis : $H_0 : \rho_{11} = 0$ vs. $H_1 : \rho_{11} \neq 0$ → if fail to reject → then prefer AR(1)

• Test statistic : $T_n = \sqrt{T} \cdot \hat{\rho}_{11} = \sqrt{200} \times 0.83 = 11.74 \sim N(0,1)$

• Critical value at $\alpha = 5\%$: $z_{\frac{\alpha}{2}}$ or 2.5% $= 1.96$

• P-value : p-value $= 2 \times Pr(z < -11.74) \approx 0\%$

* P-value (0%) < α (5%)

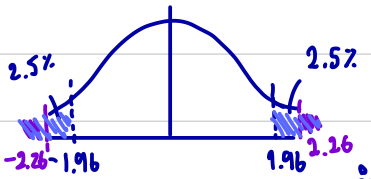
Thus, we can reject $H_0 : \rho_{11} = 0$. So, we have to do hypothesis test for AR(2).

<cont.>

VI ⓒ ▸ Hypothesis testing for AR(2) or $p = 2$?

AR(2): $y_t = \rho_{21} y_{t-1} + \rho_{22} y_{t-2} + e_t$

- Hypothesis: $H_0: \rho_{22} = 0$ vs. $H_1: \rho_{22} \neq 0$
- Test statistic: $T_n = \sqrt{T} \cdot \hat{\rho}_{22} = \sqrt{200} \times 0.16 = 2.26 \sim N(0,1)$
- Critical value at $\alpha = 5\%$: $z_{\frac{\alpha}{2}} = 1.96$
- P-value: p-value $= 2 \times Pr(z < -2.26) = 2 \times (1 - 0.9881) = 0.0238$ or 2.38%



* P-value (2.38%) < α (5%)

We can reject $H_0: \rho_{22} = 0$.

∴ AR(2) model would be more preferred than AR(1).

VII

7. A researcher uses a dataset that comprises a sample of 545 (n) full-time working males who have completed their schooling by 1980. This is a balanced panel covering the period from years 1980 to 1987. He does panel data analysis and gets the following results $T = 1987 - 1980 + 1 = 8$

Dependent Variable: ln(wage)

Regressors	Between	Fixed Effect	Pooled OLS	Random Effect
Constant	0.490	-	-0.034	-0.104
	(0.221)		(0.065)	(0.111)
Schooling	0.095	-	0.099	0.101
	(0.011)		(0.005)	(0.009)
Experience	-0.050	β_1 0.116	0.089	0.112
	(0.050)	(0.008)	(0.010)	(0.008)
Experience2	0.0051	β_2 -0.0043	-0.0028	-0.0041
	(0.0032)	(0.0006)	(0.0007)	(0.0006)
Union Member	0.274	β_3 0.081	0.180	0.106
	(0.047)	(0.019)	(0.017)	(0.018)
Married	0.145	β_4 0.045	0.108	0.063
	(0.041)	(0.018)	(0.016)	(0.017)
Black	-0.139	-	-0.144	-0.144
	(0.049)		(0.024)	(0.048)
Hispanic	0.005	-	0.016	0.020
	(0.043)		(0.021)	(0.043)
Public Sector	-0.056	β_5 0.035	0.004	0.030
	(0.109)	(0.039)	(0.037)	(0.036)

where *Wage* indicate earning in unit of dollars per month; *Schooling* is years of education, *Experience* is working experience in unit of years, *Union member* is a dummy variable taking value 1 if the worker is a member of a union; *Married* is a dummy variable taking value 1 if the worker is married; *Black* is a dummy variable taking value 1 if the worker is African American; *Hispanic* is a dummy variable taking value 1 if the worker is Hispanic; and *Public Sector* is a dummy variable taking value 1 if the worker works in the public sector. Standard errors are displayed in parentheses.

VII

a. Explain why the researcher does not present the resulting coefficients of *Schooling, Black,* and *Hispanic* estimated by the Fixed Effect estimator?

b. Suppose somebody questions that married male workers have higher wage than the others. Use the result from the Fixed Effect estimator, how would you test this hypothesis? State the null hypothesis, alternative hypothesis, test-statistics, distribution of the test statistics, 5% critical value, and the p-value.

c. Construct a 95% confidence interval for the effect of getting married on wage.

Topic 10
pg. 25

(a) Researcher does not present the resulting coefficient of Schooling, Black, and Hispanic because the value of these three regressors doesn not change across time and it remains the same (time-invariant). Thus, $X_{it} - \bar{X}_i = 0$ for all i and it is invalid as a regressor.

(b) Hypothesis testing for whether married male workers have higher wage than others: (> one-tailed test)

- Hypothesis: $H_0: \beta_4 = 0$ vs. $H_1: \beta_4 > 0$
- Test-statistics:

Using Fixed Effect Estimator:

$$T_n = \frac{\hat{\beta}_4^{FE}}{SE(\hat{\beta}_4^{FE})} = \frac{0.045}{0.018} = 2.5 \sim t_{\underbrace{545 \times (8-1) - 5}_{n(T-1) - \#\text{ of regressors}}} \approx N(0,1)$$

high degree of freedom

- Critical value at $\alpha = 5\%$: $Z_{\alpha \text{ or } 5\%} = 1.645$
- P-value $= P(Z > 2.5) = 1 - 0.9938 = 0.0062$ or $0.62\% < \alpha\ (5\%)$

$\therefore$ We can reject $H_0: \beta_4 = 0$. Thus, we can conclude that married male workers have higher wage than other.

(c) 95% confidence interval for the effect of getting married on wage: (two-tail ±)

$$95\%\ CI = \hat{\beta}_4^{FE} \pm Z_{\frac{5\%}{2}} \cdot se(\hat{\beta})$$
$$= 0.045 \pm 1.96(0.018)$$
$$= 0.045 \pm 0.03528$$

$\therefore$ 95 CI $\in (0.00972, 0.08028)$

VII

d. Suppose that the researcher carries out the Hausman Test and get the test statistic equal to 7.90. State the distribution of the Hausman Test Statistics, the critical value at 5% significance level, whether you can reject the null hypothesis, and your conclusion about Random Effect model.

e. Do you think the Between estimator is consistent if you reject the null hypothesis of the Hausman test? Why?

(d) Hausman Test is used to see if data fits RE (Random Effect Model) or FE (Fixed Effect Model) better.

- Hypothesis: $\underbrace{H_0: \beta_{RE} = \beta_{FE}}_{\text{If fail to reject} \,=\, \text{Random Effect Model}}$ vs. $\underbrace{H_1: \beta_{RE} \neq \beta_{FE}}_{\text{If reject } H_0 \,=\, \text{Fixed Effect Model}}$

- Distribution of Hausman Test Statistics is $\sim \chi^2_K$ — Chi-square distribution

  K = # of regressors = 5 from FE estimator

- Critical value for χ^2_5 at $\alpha = 5\%$ is 11.07.

∴ As test statistic of 7.9 < critical value of 11.07, we fail to reject H_0. Thus, Random Effect Model fits the data better than Fixed Effect model.

(e) If we reject null hypothesis of Hausman test, then Fixed Effect Model fits data better. In this case, Between estimator is not consistent because there is endogeneity problem when applied to Fixed Effect Model.

$\hookrightarrow \text{cov}(\alpha_i, \bar{x}_i) \neq 0$

f. Do you think the Pooled OLS estimator is consistent if you can reject the null hypothesis of the Hausman test? Why?

(f) As $\text{cov}(\alpha_i, x_{it}) \neq 0$, OLS estimator is inconsistent.

or (IX)

$\boxed{\text{VIII}}$

$$Y_{t+1} = 0.9Y_t + e_{t+1} + 0.5e_t + 0.2e_{t-1}$$

$$\begin{aligned} cov(y_t, 0.2e_{t-1}) &= 0.2\, cov(y_t, e_{t-1}) \\ &= 0.2\,[cov(0.9y_{t-1}, e_{t-1}) + cov(0.5e_{t-1}, e_{t-1})] \\ &= 0.2(0.9+0.5)\,\sigma_e^2 \end{aligned}$$

$y_{t-1} = \ldots + e_{t-1} + \ldots$

9. Let $Y_t = 0.9Y_{t-1} + e_t + 0.5e_{t-1} + 0.2e_{t-2}$, where $\{e_t\}$ is the White-Noise process. Compute the covariance $cov(Y_t, Y_{t+k})$ for $k = 1, 2, 3, 4$ in terms of variance of Y, and variance of e, i.e. σ_Y^2, σ_e^2 respecitively.

Topic 8, pg.5

* From $\{e_t\}$ is White-Noise process → $E[e_t] = 0$, $E[e_t e_s] = 0$, $E[e_t^2]$ is constant

$y_t = 0.9y_{t-1} + e_t + 0.5e_{t-1} + 0.2e_{t-2}$

- $$\begin{aligned} cov(y_t, y_{t+1}) &= cov(y_t, \overset{\rho}{0.9}y_t + e_{t+1} + \overset{\alpha}{0.5}e_t + \overset{\beta}{0.2}e_{t-1}) \\ &= cov(y_t, 0.9y_t) + \underset{\alpha}{cov(y_t, e_{t+1})} + \underset{\beta}{cov(y_t, 0.5e_t)} + \underset{\rho}{cov(y_t, 0.2e_{t-1})} \\ &= 0.9\,\sigma_y^2 + 0 + 0.5\,\sigma_e^2 + 0.2(0.9+0.5)\,\sigma_e^2 \\ &= \overset{\rho}{0.9}\,\sigma_y^2 + \overset{\alpha+\beta\rho+\beta\alpha}{0.78}\,\sigma_e^2 \end{aligned}$$

- $$\begin{aligned} cov(y_t, y_{t+2}) &= cov(y_t, 0.9y_{t+1} + e_{t+2} + 0.5e_{t+1} + 0.2e_t) \\ &= cov(y_t, 0.9y_{t+1}) + cov(y_t, 0.2e_t) \\ &= 0.9\,cov(y_t, y_{t+1}) + 0.2\,\sigma_e^2 \\ &= 0.9(0.9\sigma_y^2 + 0.78\,\sigma_e^2) + 0.2\,\sigma_e^2 \\ &= \overset{\rho^2}{0.81}\,\sigma_y^2 + \overset{\rho(\alpha+\beta\rho+\beta\alpha)+\beta = \alpha\rho+\beta\rho^2+\beta\alpha\rho+\beta}{0.902}\,\sigma_e^2 \end{aligned}$$

- $$\begin{aligned} cov(y_t, y_{t+3}) &= cov(y_t, 0.9y_{t+2} + e_{t+3} + 0.5e_{t+2} + 0.2e_{t+1}) \\ &= 0.9\,cov(y_t, y_{t+2}) \\ &= 0.9(0.81\,\sigma_y^2 + 0.902\,\sigma_e^2) \\ &= \overset{\rho^3}{0.729}\,\sigma_y^2 + \overset{\rho(\alpha\rho+\beta\rho^2+\beta\alpha\rho+\beta)}{0.8118}\,\sigma_e^2 \end{aligned}$$

- $$\begin{aligned} Cov(y_t, y_{t+4}) &= cov(y_t, 0.9y_{t+3} + e_{t+4} + 0.5e_{t+3} + 0.2e_{t+2}) \\ &= 0.9(y_t, y_{t+3}) \\ &= 0.9(0.729\,\sigma_y^2 + 0.8118\,\sigma_e^2) \\ &= \overset{\rho^4}{0.6561}\,\sigma_y^2 + \overset{\rho^2(\alpha\rho+\beta\rho^2+\beta\alpha\rho+\beta)}{0.73062}\,\sigma_e^2 \end{aligned}$$

The pattern repeats after $k > 2$

think about regressand & regressor & error term

endogeneity ??

$$y = \beta x + a + u$$

	Estimator	I Pool cross-section ($\alpha = 0$)	II Fixed effect ($cov(\alpha, x) \neq 0$)	III Random effect ($cov(\alpha, x) = 0$)
	OLS	✓	✗	✓
	FIGLS	✓	✗	✓
measurement error in x	Between	✓	✗	✓
	First diff	✓	✓	✓
	Fixed Eff	✓	✓	✓
	Random Eff	✓	✗	✓
endogeneity	2SLS/IV	✓	✓	✓
	MLE			