Take Home Midterm Exam
EE426
Semester 2/2020

There are six questions - answer all questions.
Each question is worth 20 points.
You will have your own data set. Please download your data set (`#.zip`) based on your assigned number on the list.
Time: 6 Hours (9:00 – 15:00 hr.) *(approximately one hour per question)*
(submit your answer before 3.00 pm. in ONE pdf-file by setting the file-name as No_Yourname_Student ID, e.g. 1_6104640021_ THAMCHANOK)

No.	Student ID	Name- Surname	Files	Zip-file Download
1	6104640021	Mr THAMCHANOK PIANMUEAN	Midterm_q*_1.dta	01.zip
2	6104640112	Mr HARIT HATAWONG	Midterm_q*_2.dta	02.zip
3	6104640328	Miss WIRINRATCH KIRIRAK	Midterm_q*_3.dta	03.zip
4	6104641086	Mr NATTHAPHONG TISAVIPAT	Midterm_q*_4.dta	04.zip
5	6104641300	Mr NUNTAYOD TULAKARNWONG	Midterm_q*_5.dta	05.zip
6	6104641318	Miss PHATAKAN KANCHANAPRADIST	Midterm_q*_6.dta	06.zip
7	6204640087	Mr THANAKRIT METHASATE	Midterm_q*_7.dta	07.zip
8	6204640137	Mr TANATAT BHURIVATANA	Midterm_q*_8.dta	08.zip
9	6204640178	Miss MAI SURINTRABOON	Midterm_q*_9.dta	09.zip
10	6204641218	Miss PORNWARAT FOONGPIRIYA	Midterm_q*_10.dta	10.zip
11	6204641267	Mr PONGPANOT CHALOEMVISETPHON	Midterm_q*_11.dta	11.zip
12	6204641473	Mr WIRAPHAT LIN	Midterm_q*_12.dta	12.zip
13	6204641515	Mr NARIN TANTANGTRONG	Midterm_q*_13.dta	13.zip

1. From the data set "Midterm_q1_no.dta":
   Estimate the following models

$$y_{1t} = \beta_{10} + \gamma_{12} y_{2t} + \beta_{13} x_{3t} + u_{1t} \quad (1)$$

$$y_{2t} = \beta_{20} + \gamma_{21} y_{1t} + \beta_{21} x_{1t} + \beta_{22} x_{2t} + u_{2t} \quad (2)$$

   a. Estimate model (1) and (2) using Ordinary Least Squares (OLS) and state consequences of using OLS in this case (5 Points)

   b. Estimate model (1) and (2) using Two Stage Least Squares (2SLS) and state reduced form and structural form of these simultaneous equation models. Specify endogenous variables and exogenous variables. Then, estimate reduced form models using OLS and structural form models using IV technique from the predicted endogenous variables from reduced form estimated results. (7 points)

   c. Estimate model (1) and (2) using Three Stage Least Squares (3SLS) and give the explanation of the differences among the three estimation methods (OLS, 2SLS, and 3SLS) (conceptually). Point out the advantage and disadvantage in term of properties of estimated results from each method (single equation estimation vs system equations estimation methods). (8 points)

2. From the data set "Final_q2_no.dta":

The study of cost function assumes the model follows CES cost function:

$$c = \lambda \left[\theta R^{-\alpha} + (1-\theta) W^{-\alpha} \right]^{-\frac{\beta}{\alpha}} + \varepsilon \quad (3)$$

where:

c = Total cost
R = Capital cost
W = Labor cost
λ = Efficiency Parameter
θ = Distribution Parameter
β = Parameter
α = Substitution Parameter
ε = Disturbance Term

Elasticity of Substitution can be computed from: $\sigma = \dfrac{1}{1-\theta}$

The model can then be transformed as:

$$\ln c = \ln \lambda - \frac{\beta}{\alpha} \ln \left[\theta R^{-\alpha} + (1-\theta) W^{-\alpha} \right] + \varepsilon \quad (4)$$

   a. Estimate the model (4) using NLS estimation method using initial values of $\ln\lambda$=1, θ=0.5, β=0.5, α= –0.5. Determine the estimated value of efficiency parameter (λ), distribution parameter (θ), parameter (β), and substitution parameter (α), and elasticity of substitution (σ). Perform F-test to test whether θ=0, α=0, and β=0. (6 points)

b. What will happy if we change initial values to $\ln\lambda$=0.5, θ=0.1, β=0.1, α= –0.1? Will the estimated results be the same as (a)? What are the differences between the previous result in (a) and the new result? Give explanation why? (6 points)

c. From (b), if we change convergence value from default of 0.00001 or (1e-5) to (i) 0.1 or (1e-1) and (ii) (1e-15) with maximum iteration of 40, what will happen to the estimated result? Interpret the estimated result and why do we get this kind of result? (Make comparison between previous result in (b) and the new result) (6 points)

d. Why do we prefer to estimate nonlinear regression model in log-form instead of its original functional form? (2 points)

3. From the data set "Midterm_q3_no.dta": (using do-file from assignment 4)

Index function for the decision model can be stated as.

$$I_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{5}$$

The log-likelihood function of this model is as follows:

$$\ln L = \begin{cases} \ln \Phi(z_i) \; if \;\; y_i = 1 \\ \ln \Phi(-z_i) \; if \;\; y_i = 0 \end{cases} \tag{6}$$

where: y_i = 1 for yes and 0 for no.

x_{1i} = independent variable 1 of individual *i*

x_{2i} = independent variable 2 of individual *i*

Let $\Phi(.)$ = Logistic probability distribution function. $\Phi(z_i) = \dfrac{1}{1+e^{-z_i}}$

a. Estimate the above models using MLE with (i) Newton-Ralphson algorithm; (ii) Berndt-Hall-Hall-Hausman algorithm; and (iii) Broyden-Fletcher-Goldfarb-Shanno algorithm, make comparison of the estimated results using different algorithm, and give explanation why are they different? (5 points)

b. Perform hypothesis testing whether $\beta_1 = \beta_2 = 0$ using LR-test and Wald test. Make comparison between the two tests. Which test is preferable? Why? (5 points)

c. Why overall test in MLE is Chi-square test instead of F-test? Can we still employ F-test as overall test? Why or why not? (2 points)

Let $\Phi(.)$ = Cumulation standard normal probability distribution function and

$$z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{7}$$

Assume that there exists heteroskedasticity in the model as: $\sigma_i^2 = \exp(\gamma x_{3i})^2$, then,

$\Phi(.)$ = Cumulation standard normal probability distribution function $\Phi \;\; z_i / \exp(\gamma x_{3i})$

d. Estimate the models with heteroskedasticity using MLE with Newton-Ralphson algorithm. Perform LR-test whether there exists significant heteroskedasticity.

In this case, can we perform Wald-test or LM-test to test heteroskedasticity? Why? or why not? (5 points)

e. Why individual test in MLE is *z*-test instead of *t*-test? Why don't we have R-square in MLE? If we would like to make comparison between the two estimated results, which statistical index should be employed? (3 points)

4. From the data set "Midterm_q4-1_no.dta":
The model of interest rate structure, continuous time model can be specified as:

$$r_{t+\Delta t} - r_t = (\alpha + \beta r_t)\Delta t + \varepsilon_{t+\Delta t} \quad (8)$$

where: $E[\varepsilon_{t+\Delta t}] = 0$ and $E[\varepsilon^2_{t+\Delta t}] = \sigma^2 r_t^{2\gamma} \Delta t$

Then, the model can be transformed to be discrete time model by setting $\Delta t = 1$.
The discrete time model can be stated as:

$$r_{t+1} - r_t = \Delta r_t = \alpha + \beta r_t + \varepsilon_{t+1} \quad (9)$$

where: $E[\varepsilon_{t+1}] = 0$ and $E[\varepsilon^2_{t+1}] = \sigma^2 r_t^{2\gamma}$

The model consists of four parameters, including $\alpha, \beta, \sigma^2, \gamma$. Four moment condition equations of the model can be stated as:

(1) Zero mean condition: $E(\varepsilon_{t+1}) = 0$

(2) Orthogonality condition: $E(\varepsilon_{t+1} r_t) = 0$

(3) Variance condition: $E(\varepsilon^2_{t+1} - \sigma^2 r_t^{2\gamma}) = 0$

(4) Zero covariance condition: $E\left(\left(\varepsilon^2_{t+1} - \sigma^2 r_t^{2\gamma}\right) r_t\right) = 0$

The above model can be claimed as unrestricted model for other two interest rate structure models which can be indicated as follows:

Model	α	β	σ^2	γ
(1) Unrestricted				
(2) Merton		0		0
(3) Vasicek				0

a. Estimate Unrestricted model using method of moment, Merton model using GMM, and Vasicek using GMM. Make evaluation of the estimated result of Merton model. (5 points).

b. Determine which model is the most appropriated model. Provide and give explanation of your selected criteria. (5 points)

From the data set "Midterm_q4-2_no.dta":

The model: $$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i \quad (10)$$

where: $u_i \sim N(0,\sigma^2)$, $E(x_{3i}u_i)=0$ and $E(x_{4i}u_i)=0$ but $E(x_{1i}u_i)\neq 0$ and $E(x_{2i}u_i)\neq 0$

c. What will happen to the estimated results if we estimate this model (10) using OLS? (2 points)

If $E(z_{1i}u_i)=0$, $E(z_{2i}u_i)=0$, $E(z_{3i}u_i)=0$, and z_{1i}, z_{2i}, and z_{3i} are highly correlated with x_{1i} and x_{2i}

d. Estimate the model (10) using GMM and 2SLS by employing z_{1i}, z_{2i}, and z_{3i} as instrumental variables for x_{1i} and x_{2i}. Perform the test to check whether GMM is appropriated. (5 points).

e. According to the estimated results of (d), give explanation of the differences between 2SLS and GMM estimated results in this case. (3 points)

5. From the data set "Midterm_q5_no.dta":
In the study of decision to pay dividend of listed company in the stock market, the study states the following probability function:

$$I_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \quad (11)$$

The log-likelihood function of this model is as follows:

$$\ln L = \begin{cases} \ln \Lambda(I_i) \text{ if } y_i = 1 \\ \ln \Lambda(-I_i) \text{ if } y_i = 0 \end{cases} \quad (12)$$

where: y_i = 1 for dividend paying firm and 0 otherwise.
x_{1i} = Liquidity ratio of firm *i*
x_{2i} = log of firm size firm *i*
x_{3i} = Profitability index of firm *i*

(a) Estimate the model assuming that the probability function is cumulative normal distribution function and logistic probability distribution. Can we compare the estimated coefficients of the two estimated functional forms? Why? or why not? Also, make interpret the estimated result of the Logit model (Overall test, individual test, pseudo R^2, counted R^2). (8 points)

(b) Using logistic probability distribution, compute marginal effect at mean and at median. Make interpretation of marginal effects at mean of x_{1i}. (5 points)

(c) Perform hypothesis testing whether $\beta_1 = \beta_2 = \beta_3 = 0$ using LR-test and Wald test. Perform hypothesis testing whether $\beta_1 = \beta_2$ using LR-test. Make conclusion of the tests. (5 points)

(d) Why is threshold in computing Counted R^2 important? If we change the threshold, can the value of counted R^2 change? Why? (2 points)

6. From the data set "Midterm_q6_no.dta":

Panel Data Model

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (13)$$

where: y_{it} is Dependent variable.
x_{kit} is Independent variable k.
u_{it} is Stochastic disturbance term.

Fixed Effects Model

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (14)$$

where: α_i is Cross-sectional fixed effects

Random Effects Model

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (15)$$

and $u_{it} = v_i + \varepsilon_{it}$

where: v_i = Cross-section random effects
ε_{it} = residual terms

a. Estimate model (13) using Panel Least Squares estimation method and PGLS assuming Heteroskedasticity, and test whether there exists Heteroskedasticity problem. What will happen if Heteroscedasticity occurs in the model (5 points)

b. Estimate the above three models including Panel Least Squares model (13), Fixed effects model (14), and Random-effects model (15). Perform fixed effects tests and random effects test, also state null hypothesis of the tests. Then, determine the most appropriated model. Also, give explanation of the choosing criterion (perform the tests), and make interpretation of the estimated models. (10 points)

c. What are the differences between Fixed effects estimation method and First difference estimation method? What are the differences between Fixed effects model and Random effects model? (3 points)

d. Give explanation of Within R-squares, Overall R-squares, and Between R-squares of the estimated results of the Fixed-effects model. (2 points)