Exam 4137 – Spring 2019

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Your answers should be as short as possible and as long as necessary. Total marks of the examination is 140. Each subquestion in problem 1 carries 5 marks. Sub-questions in problem 2 and 3 carry 10 marks each.

---

1. Give brief answers to the following question.

   (a) True or false: in case of heteroscedasticity, both OLS and WLS will give consistent estimates, but OLS is less efficient and its standard error formula is wrong.

   ANSWER HINT: True.

   (b) Suppose the model of interest is

   $$y_i = \alpha + \beta x_i + \gamma z_i + \varepsilon_i$$

   where $E[\varepsilon_i|x_i,z_i] = 0$, $Var(\varepsilon_i|x_i,z_i) = \sigma^2$, and $cov(x_i,z_i) = 0$. Will the bivariate regression of Y_i on X_i have the same coefficient estimate and standard error for as the multivariate regression of Y_i on X_i and Z_i?

   ANSWER HINT:
   The parameter estimate will be the same. But the standard error will be bigger in the bivariate case.

   (c) Consider the case where there is one endogenous regressor and one instrument, what is meant by suggesting that an instrument is weak? And what is the rule of thumb used to check whether an IV is weak or not?

   ANSWER HINT:
   An instrument is weak if it has a small(weak) relationship with the endogenous regressor. This can be directly checked by looking at the F-statistic (>10) on the instrument in the first stage.

   (d) Consider the panel data model $y_{it} = \beta_0 + \beta_1 x_{it} + a_i + \varepsilon_{it}$, $i = 1,2,..,N$, $t = 1,2,..,T$, where x_{it} is strictly exogenous. Suppose that $E[a_i|x_{it}] \neq 0$. Explain how you will estimate the parameter β_1.

   ANSWER HINT:
   Fixed effect model: can use either Within estimator (demeaned regression) or FD estimator.

2. We will look at the student choice of an on-campus eating place at lunch time. Students have three different options, The Deli (serving sandwiches, soups and salads), Quik Stop (sandwiches, pizza by slice, hot dog etc) and Pacific Corner (sushi, stir fry, noodles and curries). Both The Deli and Quik Stop offer more western cuisine, while Pacific Corner offers largely Asian fare. The lunches are

comparable in price across all three places.To attract the price-sensitive students, the three restaurants often make special promotions by cutting the price of lunch by 5 kroner. We are interested in the question how effective these promotions are.

We have collected information on 200 student choice during the course of a semester. We also have information on the availability of promotions from each restaurant at the time of choice.

sid	student ID
rid	Restaurant ID (1=Deli, 2=Quik, 3=Pacific)
chosen	choice indicator (1 if chosen, 0 otherwise)
promotion	Promotion indicator (1 if having promotion, 0 otherwise)

If the student ID (x_i) assignment is random, we should expect that x_i to be independent from the student's preference. We starts out by studying whether the student ID has any effect on the student's preference between Western cuisine and Asian fare:

(a) Specify a logit model for this purpose. Write down the log-likelihood functions. And outline how one can estimate the model using MLE.

ANSWER HINT:
Define $y_i = 1$,if individual chooses Pacific Corner.
Logit: $P(y_i = 1|x_i) = E[(y_i = 1|x_i)] = \frac{exp(\alpha+\beta x_i)}{1+exp(\alpha+\beta x_i)}$
log-likelihood function: $\log L = \sum_{i=1}^{N}[y_i \log P(y_i = 1|x_i) + (1-y_i)\log P(y_i = 0|x_i)]$
write down log-likelihood, and maximize. ie numerically iterate until first order conditions evaluated at the parameter values are zero, and second derivative is negative. the main idea is that there are no other parameters under which our data are more likely to occur.

(b) Based on the attached Stata output, do you think that the student ID has any impact on the preference between Western cuisine and Asian fare?

ANSWER HINT:
Negative and statistically significant at 5% level. Students with higher Student ID numbers have a lower probability choosing Asian fare than students with lower ID numbers. May indicate that the student ID assignment is not purely random.

Suppose that we want study the effect of promotions on choices of restaurants. Note that in the following you don't need to include the variable "student ID" in your model.

(c) Specify a model so that you can calculate the expected market share changes due to promotions. Explain how you can use the model to obtain, for example, the Pacific Corner's market gain if it makes a promotion.

ANSWER HINT:
We want to study the probability of choosing a restaurant as a function of the promotion dummy and restaurant dummies
Denote P_{ij} as promotion, define the utility of choosing restaurant j for student i as

$$U_{ij} = \alpha_j + \beta P_{ij} + \varepsilon_{ij}$$

where ε_{ij} i.i.d extreme value distributed. Then we have

$$Pr(chosen = j|P_{ij}) = \frac{\exp(\beta P_{ij} + \alpha_j)}{\sum_{k=1}^{3} \exp(\beta P_{ik} + \alpha_k)}$$

The market gain for a given restaurant, for example, $j = 3$, Pacific Corner is given by

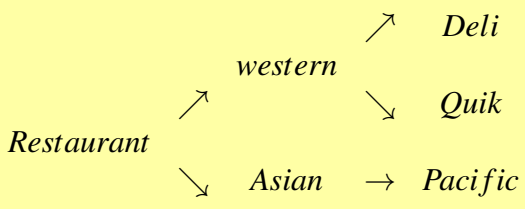
$$\begin{aligned} &Pr(chosen = j|P_{ij} = (0,0,1)) - Pr(chosen = j|P_{ij} = (0,0,0) \\ &= \frac{\exp(\alpha_3 + \beta)}{\exp(\alpha_1) + \exp(\alpha_2) + \exp(\alpha_3 + \beta)} - \frac{\exp(\alpha_3)}{\sum_{k=1}^{3} \exp(\alpha_k)} \end{aligned}$$

(d) Based on the attached Stata output, do you think promotion is an effective policy? Calculate the expected market share gain by Pacific Corner if it makes a promotion while the other two do not.

ANSWER HINT:
Yes, the estimate for promotion is positive and significantly different from zero.
The market share gain is $0.407 - 0.298 \approx 0.109$

(e) Explain briefly what is meant by the term IIA? Do you think that IIA will hold in this case? If not, can you think of a better model? You are not supposed to specify the whole model, but only to draw a decision tree .

ANSWER HINT:
IIA: Independence from Irrelevant Alternatives: ,the relative probability of someone choosing between two options is independent of any additional alternatives in the choice set.
No, the IIA might not be valid – as mentioned above The Deli and Quik Stop are closer substitutes since both of them serve western cuisine.
Nest logit would be a better model here:



3. A recent paper[1] investigates whether reality TV shows induce real effects by study the effect of the MTV program *16 and Pregnant* on the teen birth rates in the U.S. since it began broadcasting in 2009. The paper uses a panel dataset on Designated Market Area (DMA) i and estimates the following equation:

$$\log BR_{it} = \beta x_{it} + \mu_t + \gamma_i + \varepsilon_{it} \tag{1}$$

where, $\log BR_{it}$ is the log of teen birth rate by quarter of conception t in DMA i. $x_{it} = Rate_j \times Post_t$ and $Rate_j$ is average Nielsen rating (a measure of the audience size) in DMA j for the show *16 and Pregnant* for 12 to 24 year old views for the years 2009-2012. $Post_t$ is a dummy variable which equals to 1 for the quarters after the introduction of the show, and 0 otherwise. μ_t, γ_i are time and DMA fixed effects.

(a) What is the parameter of interest? What variation in the data did the authors use to estimate it in (1)?

ANSWER HINT:
The parameter of interest is β. The identification strategy used here is a comparison of teen birth rates before and after the introduction of the show, as well as the regional differences on the popularity of show.

(b) The authors are worried about the possibility that locations in which the show is more popular are not randomly selected in terms of teen childbearing rates. If this is the case, what can you say about the estimate from (1)?

ANSWER HINT:
If this is the case, then there is unobserved variable that impact both the popularity of the show x_{it} and the teen birth rate $logBR_{it}$, namely, $E(\varepsilon_{it}|x_{it}) \neq 0$. We have an endogeneity bias, OLS estimator $\hat{\beta}$ will be inconsistent.

(c) To cope with this potential problem, the authors suggested to use the information on the rating of all MTV shows before the introduction of the show *16 and Pregnant* in DMA i ($RMTV_j$) and construct an IV z_{it} for x_{it}:

$$z_{it} = RMTV_j \times Post_t$$

Discuss whether z_{it} is a valid instrument.

[1] Kearney, Melissa S., and Phillip B. Levine. "Media Influences on Social Outcomes: The Impact of MTV's "16 and Pregnant" on Teen Childbearing." The American Economic Review 105, no. 12 (2015): 3597-632. http://www.jstor.org/stable/43821387.

ANSWER HINT:
Random assignment: the rating of all MTV shows is exogenous This implies that both the reduced form and first stage regression have a casual interpretation. This will be violated if there is some unobserved factors that affect both rating and the (change) in teen birth rates. Here we may expect the students to relate their discussions to the Omitted Variable Bias.
The exclusion restriction: Conditional on the *16 and Pregnant rating,* the rating of all MTV shows does not impact the (change) in teen birth rates. In other words, all possible impacts of the instrument on the outcome go through the endogenous variable.
The relevant assumption: the rating of all MTV shows affects the rating of *16 and Pregnant rating.* First step regression can be used to check this.
The Monotonicity assumption: an increase in MTV rating should not decrease the rating of the show *16 and Pregnant.* This assumption is likely to hold.

(d) The estimates of different specifications are given in the following Table. Interpret OLS and IV estimates of β.

Table 1: Estimates of the Impact of *16 and Pregnant* Ratings on Teen Birth Rates

	OLS (1)	First Stage (2)	IV (3)	Reduced Form (4)
Dependent Variable	ln(birth rate)	*16 and Pregnant* Ratings	ln(birth rate)	ln(birth rate)
16 and Pregnant Ratings	-1.020*		-2.368**	
	(0.552)		(0.942)	
MTV Ratings 2008-09		1.511***		-3.581**
		(0.204)		(1.512)
Unemployment Rate	-1.440***	-0.001	-1.487***	-1.485***
	(0.401)	(0.026)	(0.375)	(0.432)
F-Statistic on Omitted Instrument		48.1		

Notes: The birth data used for this analysis represents quarterly birth rates by DMA for conceptions leading to live births between 2005 and 2010. The sample size in each model is 4919 (205 DMAs, 24 quarters, and one observation was dropped because there were no teen births). Coefficients and standard errors (reported in parentheses) in birth rate regressions are multiplied by 100. Each model also includes the percentage of a DMA's female teen that are Hispanic and non-Hispanic black along with quarter and DMA*season fixed effects. Regressions are weighted by the relevant sample sizes for each outcome. Reported standard errors are clustered at the DMA level.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

ANSWER HINT:
The OLS estimate is negative. Watching the show has a negative impact to the teen birth rate. Rating increases by 1 unit will lead to the teen birth rate drop by -1.0% (in relative terms). The estimate is statistically significant at 10% level.
The IV estimate is negative and statistically significant at 5 percent level, despite larger standard errors than OLS. The IV estimate is larger in magnitude than the OLS estimate: Rating increases by 1 unit will lead to the teen birth rate drop by -2.4% (in relative terms)

In the following, we study a simplified version of the model. Instead of continuous endogenous and instrument variables x_{it} and z_{it}, we use dummy variables generated from these two variables. Namely, for a given constant $c > 0$, we define $D_{it} = 1(x_{it} > c)$ and $S_{it} = 1(z_{it} > c)$. We want to estimate

$$\log BR_{it} = \beta D_{it} + \mu_t + \gamma_i + \varepsilon_{it} \tag{2}$$

using the variableS_{it} as the instrument for D_{it}.

(e) Explain how you can obtain the IV estimator using Indirect Least Squares. Note that you are supposed to specify both the first stage and reduced form equations.

ANSWER HINT:
The first step regression:

$$D_{it} = \lambda S_{it} + \mu_t + \gamma_i + u_{it} \tag{3}$$

The reduced form:

$$\log BR_{it} = \gamma S_{it} + \mu_t + \gamma_i + u_{it} \tag{4}$$

The IV estimator can be obtained by

$$\hat{\beta} = \frac{\hat{\gamma}}{\hat{\lambda}}$$

(f) Show that the reduced form equation is essentially a difference-in-differences setup.

ANSWER HINT: .
Consider the case when $T = 2$, where t=1 the period before the show introduction and t=2 is the period after the show introduction.
Control group: those with $S_{i2} = 0$
Treatment group: those with $S_{i2} = 1$
Control group $t = 1$, $E[\log BR_{it}|S_{i2} = 0, t = 1] = \mu_1 + E[\gamma_i|S_{i2} = 0]$
Control group $t = 2$, $E[\log BR_{it}|S_{i2} = 0, t = 2] = \mu_2 + E[\gamma_i|S_{i2} = 0]$
Treatment group $t = 1$, $E[\log BR_{it}|S_{i2} = 1, t = 1] = \mu_1 + E[\gamma_i|S_{i2} = 1]$
Treatment group $t = 2$, $E[\log BR_{it}|S_{i2} = 1, t = 2] = \mu_2 + \tau + E[\gamma_i|S_{i2} = 1]$
So the $\tau = DID$

(g) Based on results from (f), what do you think is the key assumption to obtain a consistent estimator for the casual effect of S_{it} on $\log BR_{it}$. How will you check this assumption? We assume that there were two periods before the introduction of the show.

ANSWER HINT: .
Common trend assumption.
We can use placebo treatment test, let $t = 0$ to be the period when the show is introduced, define the placebo treatment at $t = -1$ as $K_{it} = (RMTV_i > c) \cdot 1(t = -1)$, and run the regression

$$\log BR_{it} = \gamma_{-1} K_{it} + \gamma S_{it} + \mu_t + \gamma_i + u_{it}$$

and check whether $\hat{\gamma}_{-1}$ statistically significant different from zero.

Stata Output

```
. su sid rid promotion chosen

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         sid |        600       100.5    57.78248          1        200
         rid |        600           2    .8171778          1          3
   promotion |        600        .325    .4687657          0          1
      chosen |        600    .3333333    .4717978          0          1

. preserve
. keep if rid==3 (400 observations deleted)
. logit chosen sid

Logistic regression                             Number of obs     =        200
                                                LR chi2(1)        =       4.53
                                                Prob > chi2       =     0.0332
Log likelihood = -119.04619                     Pseudo R2         =     0.0187
------------------------------------------------------------------------------
      chosen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sid |  -.0057945   .0027569    -2.10   0.036    -.0111979   -.0003911
       _cons |  -.3116672   .2998073    -1.04   0.299    -.8992787    .2759443
------------------------------------------------------------------------------

. restore
. asclogit chosen promotion , case(sid) alternative(rid)

Alternative-specific conditional logit          Number of obs      =        600
Case variable: sid                              Number of cases    =        200
Alternative variable: rid                       Alts per case: min =          3
                                                               avg =        3.0
                                                               max =          3

                                                  Wald chi2(1)    =       7.09
Log likelihood = -215.41497                       Prob > chi2     =     0.0077
------------------------------------------------------------------------------
      chosen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rid          |
   promotion |   .4785491   .1796652     2.66   0.008     .1264118    .8306864
-------------+----------------------------------------------------------------
The_Deli     |
 (base alternative)
-------------+----------------------------------------------------------------
Quik_Stop    |
       _cons |  -.0806789   .1702172    -0.47   0.636    -.4142985    .2529406
-------------+----------------------------------------------------------------
Pacific_Co~r |
       _cons |  -.2029544   .1766092    -1.15   0.250     -.549102    .1431932
------------------------------------------------------------------------------

. predict p_ml, pr
. l sid rid chosen promotion p_ml if sid<=9, sep(3)
     +-------------------------------------------------------+
     | sid              rid   chosen   promot~n       p_ml |
     |-------------------------------------------------------|
  1. |   1         The Deli        1          0   .3086607 |
  2. |   1        Quik Stop        0          0   .2847363 |
  3. |   1   Pacific Corner        0          1    .406603 |
     |-------------------------------------------------------|
  4. |   2         The Deli        0          1   .4813464 |
  5. |   2        Quik Stop        1          0   .2751617 |
  6. |   2   Pacific Corner        0          0   .2434919 |
     |-------------------------------------------------------|
  7. |   3         The Deli        0          1   .4187665 |
  8. |   3        Quik Stop        0          0   .2393879 |
  9. |   3   Pacific Corner        1          1   .3418455 |
     |-------------------------------------------------------|
 10. |   4         The Deli        0          1   .3651227 |
 11. |   4        Quik Stop        0          1    .336822 |
 12. |   4   Pacific Corner        1          1   .2980553 |
     |-------------------------------------------------------|
 13. |   5         The Deli        1          1    .411803 |
 14. |   5        Quik Stop        0          1   .3798841 |
 15. |   5   Pacific Corner        0          0   .2083129 |
     |-------------------------------------------------------|
 16. |   6         The Deli        0          1   .4813464 |
 17. |   6        Quik Stop        0          0   .2751617 |
 18. |   6   Pacific Corner        1          0   .2434919 |
     |-------------------------------------------------------|
 19. |   7         The Deli        0          0    .302575 |
 20. |   7        Quik Stop        1          1   .4504284 |
 21. |   7   Pacific Corner        0          0   .2469966 |
     |-------------------------------------------------------|
 22. |   8         The Deli        1          0    .302575 |
 23. |   8        Quik Stop        0          1   .4504284 |
 24. |   8   Pacific Corner        0          0   .2469966 |
     |-------------------------------------------------------|
 25. |   9         The Deli        0          0   .3651227 |
 26. |   9        Quik Stop        1          0    .336822 |
 27. |   9   Pacific Corner        0          0   .2980553 |
     +-------------------------------------------------------+
```