Problem Set 3.

① . $y_1 = \gamma_{12} y_2 + \gamma_{13} y_3 + \delta_{11} z_1 + \delta_{13} z_3 + u_1$ (1)

$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + u_2$ (2)

$y_3 = \delta_{31} z_1 + \delta_{32} z_2 + \delta_{33} z_3 + \delta_{34} z_4 + u_3$ (3)

For eqn (1) $B_1 = (\gamma_{11}\ \gamma_{12}\ \gamma_{13}\ \delta_{11}\ \delta_{12}\ \delta_{13}\ \delta_{14})'_{7\times 1}$ Note $G=3,\ M=4$

by normalization; $B_1 = (-1\ \gamma_{12}\ \gamma_{13}\ \delta_{11}\ \delta_{12}\ \delta_{13}\ \delta_{14})'$; restrictions on (1) are $\delta_{12}=0$ and $\delta_{14}=0$

$\therefore$ we have $J_1 = 2$ and $G-1 = 3-1 = 2$. ; since $J_1 \geq G-1$ order condition is satisfied.;

equation (1) might be identified i.e. we need to check further, the rank condition.

Define restriction matrix $R_{1\,(J_1 \times (G+M))}$ such that $R_1 B_1 = 0$.

$$\therefore R_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_1 = (-1\ \gamma_{12}\ \gamma_{13}\ \delta_{11}\ \delta_{12}\ \delta_{13}\ \delta_{14})'$$

and
$$B = \begin{bmatrix} -1 & \gamma_{21} & \gamma_{31} \\ \gamma_{12} & -1 & \gamma_{32} \\ \gamma_{13} & \gamma_{23} & -1 \\ \delta_{11} & \delta_{21} & \delta_{31} \\ \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{13} & \delta_{23} & \delta_{33} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{bmatrix} ; \quad R_1 B = \begin{bmatrix} \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \delta_{32} \\ 0 & 0 & \delta_{34} \end{bmatrix}.$$

$\therefore$ Rank $(R_1 B) = 1 \neq G-1$; Rank condition fails.

$\therefore$ equation (1) is unidentified

For eqn (2) $B_2 = (\gamma_{21}\ \gamma_{22}\ \gamma_{23}\ \delta_{21}\ \delta_{22}\ \delta_{23}\ \delta_{24})'$

by normalization; $B_2 = (\gamma_{21}\ -1\ \gamma_{23}\ \delta_{21}\ \delta_{22}\ \delta_{23}\ \delta_{24})$ & restrictions on (2) are $\gamma_{23}=0$, $\delta_{22}=0$, $\delta_{23}=0$, $\delta_{24}=0$ $\therefore$ we have $J_2 = 4$ and $G-1 = 3-1 = 2$; since $J_2 \geq G-1$ order cond. is satisfied, equation (2) might be identified. Further check on rank cond. is required.

Define restriction matrix $R_{2\,(J_2 \times (G+M))}$ such that $R_2 B_2 = 0$.

$$\therefore R_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{4\times 7} \text{ and } B = \begin{bmatrix} -1 & \gamma_{21} & \gamma_{31} \\ \gamma_{12} & -1 & \gamma_{32} \\ \gamma_{13} & \gamma_{23} & -1 \\ \delta_{11} & \delta_{21} & \delta_{31} \\ \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{13} & \delta_{23} & \delta_{33} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{bmatrix}_{7\times 3}$$

$$\therefore R_2 B = \begin{bmatrix} \gamma_{13} & \gamma_{23} & -1 \\ \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{13} & \delta_{23} & \delta_{33} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{bmatrix} = \begin{bmatrix} \gamma_{13} & 0 & -1 \\ 0 & 0 & \delta_{32} \\ \delta_{13} & 0 & \delta_{33} \\ 0 & 0 & \delta_{34} \end{bmatrix}$$

$\therefore$ Rank $(R_2 B) = 2 = G-1$; Rank condition is satisfied and since $J_2 = 4 > G-1$, eqn (2) is overidentified

For eqn (2) variable(s) that can be used for instrumenting are/is those exogeneous variable(s) that help explain endogeneous variable y_1 ~~but excluded~~ i.e. z_1 and z_3.

(21) Consider equation (4)

First, we check order condition. In order to pass order condition, there must be at least one exogenous variables is exclude from the equation (4). Since, log(land) is excluded from equation (4), therefore, equation (4) pass order condition.

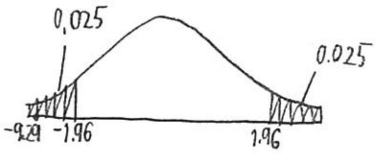
Second, we check rank condition. In order to pass rank condition, at least one of exogenous variables excluded from the equation (4) must have a nonzero population coefficient. We regress open on all exogenous variables,

$$\underset{(15.8483)}{\text{open} = 117.0845} + \underset{(1.4932)}{0.5465} \log(pcinc) - \underset{(0.8142)}{7.5671} \log(land) \quad (a)$$

$H_0: \beta_{\log(land)} = 0$

$H_a: \beta_{\log(land)} \neq 0$

$$t = \frac{-7.5671}{0.8142} = -9.29$$



Since, the computed t-value falls into rejection region. We reject H_0 at 0.05 significance level. That is, coefficient log(land) is not equal to zero. So, equation (4) pass rank condition.

In conclusion, equation (4) is identified.

Consider equation (5)

First, we check order condition. Since, there is no exogenous variable which excludes from equation (5), equation isn't identified.

Next, we want to check whether log(land) can be used as an IV for open

If we assume that log(land) doesn't correlate with error term in equation (a) and cov(open, log(land)) is not equal to zero, we can conclude that log(land) can be used as IV for open

We regress equation (4) by using log(land) as IV for open, we get

$$\underset{(15.7972)}{\text{inf} = 26.8993} - \underset{(0.1422)}{0.3375}\,\text{open} + \underset{(1.988)}{0.3758} \log(pcinc)$$

2.1) check whether eqn (4) is identified.; examin eqn (5)

<alternative>
<additional>

ORDER CONDITION : log(land) ; there's at least one exo var excluded from eqn(4)

Rank condition : $\beta_{22} \neq 0$ i.e. need to check if there's at least one of the exo excluded from eqn (4) and have nonzero pop. coefficient in eqn (5)

To check Rank condition, we obtain the reduce form for open :

$$\text{open} = \beta_{20} + \alpha_2(\beta_{10} + \alpha_1 \text{open} + \beta_{11}\log(pinc) + u_1) + \beta_{12}\log(pcinc) + \beta_{22}\log(land) + u_2$$

$$(1-\alpha_2\alpha_1)\text{open} = (\beta_{20} + \alpha_2\beta_{10}) + (\alpha_2\beta_{11} + \beta_{12})\log(pinc) + \beta_{22}\log(land) + (\alpha_2 u_1 + u_2)$$

Reduced Form. $\Rightarrow$

$$\therefore \text{open} = \underbrace{\frac{\beta_{20} + \alpha_2\beta_{10}}{1-\alpha_2\alpha_1}}_{\pi_0} + \underbrace{\frac{\alpha_2\beta_{11} + \beta_{12}}{1-\alpha_2\alpha_1}}_{\pi_1}\log(pinc) + \underbrace{\frac{\beta_{22}}{1-\alpha_2\alpha_1}}_{\pi_2}\log(land) + \underbrace{\frac{(\alpha_2 u_1 + u_2)}{(1-\alpha_2\alpha_1)}}_{v_1}$$

and test for π_2 i.e. $H_0: \pi_2 = 0$; $H_a: \pi_2 \neq 0$ since ~~test~~ t-test statistic for $\hat{\pi}_2$ is -9.29 (result attached) and P-value = .000 , we can conclude that π_2 is statistically significant at any degree of confident level. Thereby Rank condition is also satisfied. Note that with π_2 statistically significant, log(land) can ~~also~~ be used as an iv for open. endogen. var, open, with assumption that $cov(\log(land), u_1) = 0$.

2.2) OLS: $\hat{inf} = 25.234 - 0.215\, open$
$(4.102) \quad (.093)$

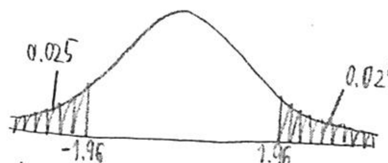
IV: $\hat{inf} = 29.607 - .333\, open$
$(5.61) \quad (.139)$

From stata output reported in table 3, we can see that after dropping lnpcinc (which is not statistically significant in IV model in 2.1 (p-value = .85, t-statistic = .19)), coefficient of open is still statistically significant at 0.05 significance level but insignificant at 0.01 level for both IV and OLS. However, when we consider the change in magnitude, OLS and IV estimators give us quite different result. Using IV, we have $\hat{\alpha}_1 = -.333$ indicating that if import as % of GDP increases by 1 percentage point, we will expect an average decrease of annual inflation of .333 percentage point ceteris paribus. This result is quite similar to $\hat{\alpha}_1 = -.337$ produced by IV before we drop lnpcinc. This is not surprising as we drop a insignificant exogeneous variable from the model which should not affects our result significantly. However, when we use OLS to estimate the model instead, we have $\hat{\alpha}_1 = -.215$; 1 percentage point increase in import as % of GDP leads to .215 percentage point decrease in annual inflation, on average, ceteris paribus. Magnitude of our estimates changes quite a lot signifying that endogeneity problem may be present in OLS model making our OLS estimator inconsistent (We need to conduct endogeneity test to confirm this). Our conclusions regarding statistical significance of our coefficient remain unchanged, however, probably because standard errors in IV are higher than in OLS.

(2.3) $inf = 24.0089 - 0.3370\, open + 0.8033 \log(pcinc) - 6.5557\, oil$
$(15.75965) \quad (0.1419) \quad (2.0804) \quad (9.6279)$

$H_0 : \beta_{oil} = 0$
$H_a : \beta_{oil} \neq 0$

$$t = \frac{-6.5557}{9.6279} = -0.68$$



Since, the computed t-value doesn't fall into rejection region. We don't reject H_0 at 0.05 significance level. That is, being an oil producer has no effect on inflation, ceteris paribus.

alternative: The computed t-statistic does not fall into rejection region. We cannot reject H_0 at 0.1 level of significance. There is not enough evidence to conclude that being an oil producer has a statistically significant partial effect on inflation rate, holding other explanatory variable fixed. It does not have ceteris paribus effect on inflation at any resonable significant level)

(3) 3.1 Since each equation has a behaviral and the decisions come from 2 agents. As equation (6), the income of each person that the employer will give depends on cigs, educ and age and equation (7), the consumption on cigarattes depends on the decision of each person, determined by log(income), educ, age, log(cigpric), restaurn. Since the decisions come from 2 agents then this is "SEM" and then we can ceteris paribus interpret each equation.

we expect that the sign of γ_5 is <u>negative</u> because the increase in a pack of cigarette price should reduce the consumption on cigarette perday.

We expect that the sign of γ_6 is <u>negative</u> because the person lives in a state with restaurant smoking restriction should smoke <u>less</u> than who not lives in a state with restaurant restriction.

✓

3.2) (1) Order condition: there are 2 exogeneous variable excluded from eq(6) which are "log(cigpric)" and "restaurn" ✓

(2) Rank condition: "log(cigpric)" and "restaurn" are statistically significant in eq(7)

∴ Assume $\gamma_5 \neq 0$ and or or $\gamma_6 \neq 0$, and $Cov(\log(cigpric), u_1) = 0$, $Cov(restaurn, u_1) = 0$ ✓

(1), (2), and the assumptions stated are satisfied, Eq(6) is identified

3.3) $cigs = \gamma_0 + \gamma_1(\beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 age + \beta_4 age^2 + u_1) + \gamma_2 educ + \gamma_3 age + \gamma_4 age^2 + \gamma_5 \ln cigprice + \gamma_6 restau$

$\therefore\ cigs = \dfrac{\gamma_0 + \gamma_1\beta_0}{1-\gamma_1\beta_1} + \dfrac{\gamma_1\beta_2 + \gamma_2}{1-\gamma_1\beta_1} educ + \dfrac{\gamma_1\beta_3 + \gamma_3}{1-\gamma_1\beta_1} age + \dfrac{\gamma_1\beta_4 + \gamma_4}{1-\gamma_1\beta_1} age^2 + \dfrac{\gamma_5}{1-\gamma_1\beta_1} \ln cigprice + \dfrac{\gamma_6}{1-\gamma_1\beta_1} restaurn$

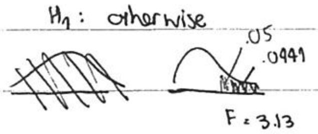
$+ \dfrac{\gamma_1 u_1 + u_2}{1-\gamma_1\beta_2}$

Reduced form: $cigs = \pi_0 + \pi_1 educ + \pi_2 age + \pi_3 age^2 + \pi_4 \ln cigprice + \pi_5 restaurn + v$ ✓

$\hat{cigs} = 1.58 - .45\, educ + .823\, age - .0096\, age^2 - .351 \ln cigprice - 2.736\, restaurn$ ✓

$H_0: \pi_4 = \pi_5 = 0$

H_1: otherwise



According to stata output, F-statistic = 3.13 falls into rejection region. We reject $H_0: \pi_4 = \pi_5 = 0$ at 0.05 level of significance and conclude that lncigprice and restaurn can jointly statistically determine cigs, holding other explanatory variables fixed. That is rank condition holds and we can identify (6).

lncigprice and restaurn can be used as instruments for cigs. ✓ However, because lncigprice is individually not statistically significant at any conventional significant level (p-value = .951), it cannot be used as instrument alone ✓ (When we run reduced form regression but drop restaurant, lncigprice is still statistically insignificant at any conventional level and dropping lncigprice from reduced form equation change the result only slightly.) Individually, restaurn can be used as IV for cigs while lncigprice cannot, assuming that both are exogeneous. ✓

good!

(lncigprice and restaurn as IVs for cigs)

3.4) $\widehat{lnincome} = 7.781 - 0.042\,cigs + 0.0397\,educ + 0.0938\,age - 0.001\,age^2$ ~~-0.00~~
$\qquad\qquad\quad (.229) \quad (.026) \quad (.016) \quad (.024) \quad (.0003)$

(Note that we get very similar result when using only restaurn as IV for cigs)

3.5) OLS: $\widehat{lnincome} = 7.795 + 0.0017\,cigs + 0.06\,educ + 0.058\,age - .0006\,age^2$
$\qquad\qquad\quad (.17) \quad (.0017) \quad (.0079) \quad (.0076) \quad (.000)$

OLS, $\hat{\beta}_1^{OLS} = 0.0017$ means that if average number of cigarettes smoked per day increases by 1, income is expected to increase by 0.17%, on average, holding all other explanatory variable fixed.
This implies that cigs has a positive effect on income holding other explanatory variable fixed. which is quite not what we expect. However, we cannot reject $H_0: \beta_1 = 0$ at any conventional significant level (p-value = 0.313)
Therefore, we can at best conclude that ~~the~~ there is not enough evidence that number of ciggarettes smoked per day has an ceteris paribus effect on income.
When we use IV estimator, $\hat{\beta}_1^{IV} = -.042$ implying that, holding other factors constant if individual increases his/her ciggarette consumption per day by 1 (on average), his/her income is expected to fall by average of 4.2%. Although this coefficient estimate is also statistically insignificant at 0.1 significant level (p-value = 0.107) ~~it~~ it has expected sign.

Standard error of $\hat{\beta}_1$ (and also of other β) are much larger when we use IV than when we use OLS. This can be one of the reasons why our IV estimates for β_1 is still statistically insignificant eventhough it is much larger (in absolute term) when we compare it ~~the~~ with OLS estimate.
Because OLS estimate ~~has~~ for β_1 has unexpected sign, we suspect that we are facing endogeneity problem when using OLS and our OLS estimator suffers from asympotic bias

Question 2.1

Table1

```
. reg open lpcinc lland
```

Source	SS	df	MS
Model	28606.1936	2	14303.0968
Residual	35151.7966	111	316.682852
Total	63757.9902	113	564.230002

Number of obs	=	114
F(2, 111)	=	45.17
Prob > F	=	0.0000
R-squared	=	0.4487
Adj R-squared	=	0.4387
Root MSE	=	17.796

open	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lpcinc	.5464812	1.49324	0.37	0.715	-2.412473	3.505435
lland	-7.567103	.8142162	-9.29	0.000	-9.180527	-5.953679
_cons	117.0845	15.8483	7.39	0.000	85.68005	148.489

Table2

```
. ivregress 2sls inf (open=lland) lpcinc
```

Instrumental variables (2SLS) regression

Number of obs	=	114
Wald chi2(2)	=	5.73
Prob > chi2	=	0.0570
R-squared	=	0.0309
Root MSE	=	23.52

inf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
open	-.3374871	.1422122	-2.37	0.018	-.6162179	-.0587563
lpcinc	.3758247	1.98839	0.19	0.850	-3.521348	4.272997
_cons	26.89934	15.1972	1.77	0.077	-2.886624	56.6853

Instrumented: open
Instruments: lpcinc lland

Question 2.2

Table3

	(1) inf b/se	(2) inf b/se	(3) inf b/se
open	-0.3375**	-0.2150**	-0.3329**
	(0.142)	(0.093)	(0.139)
lpcinc	0.3758		
	(1.988)		
_cons	26.8993*	25.2342***	29.6066***
	(15.197)	(4.102)	(5.608)
r2	0.031	0.045	0.032
N	114	114	114

* p<0.1, ** p<0.05, *** p<0.01

Question 2.3

Table4

```
. ivregress 2sls inf (open=lland) lpcinc oil

Instrumental variables (2SLS) regression          Number of obs  =       114
                                                   Wald chi2(3)   =      6.24
                                                   Prob > chi2    =    0.1006
                                                   R-squared      =    0.0349
                                                   Root MSE       =    23.471
```

inf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
open	-.3369707	.1419215	-2.37	0.018	-.6151316	-.0588097
lpcinc	.8032896	2.080441	0.39	0.699	-3.274299	4.880878
oil	-6.555731	9.627909	-0.68	0.496	-25.42609	12.31462
_cons	24.00886	15.75165	1.52	0.127	-6.863805	54.88153

```
Instrumented:  open
Instruments:   lpcinc oil lland
```

Question 3.3

Table 5

```
. reg cigs educ age agesq lcigpric restaurn
```

Source	SS	df	MS
Model	7740.15214	5	1548.03043
Residual	144013.531	801	179.792173
Total	151753.683	806	188.280003

```
Number of obs  =       807
F(  5,   801)  =      8.61
Prob > F       =    0.0000
R-squared      =    0.0510
Adj R-squared  =    0.0451
Root MSE       =    13.409
```

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.4501466	.1616396	-2.78	0.005	-.7674338	-.1328594
age	.822541	.1543224	5.33	0.000	.5196168	1.125465
agesq	-.0095903	.0016792	-5.71	0.000	-.0128864	-.0062942
lcigpric	-.3513161	5.76555	-0.06	0.951	-11.66869	10.96606
restaurn	-2.736389	1.109693	-2.47	0.014	-4.914639	-.5581394
_cons	1.580112	23.69558	0.07	0.947	-44.93266	48.09289

test lcigpric restaurn

(1) lcigpric = 0
(2) restaurn = 0

F(2, 801) = 3.13
Prob > F = 0.0441

Question 3.4

Table 6

```
. ivregress 2sls lincome (cigs=lcigpric restaurn) educ age agesq
```

Instrumental variables (2SLS) regression

Number of obs =	807
Wald chi2(4) =	89.80
Prob > chi2 =	0.0000
R-squared =	.
Root MSE =	.87723

lincome	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
cigs	-.0421257	.0261371	-1.61	0.107	-.0933535	.009102
educ	.0396746	.0162305	2.44	0.015	.0078633	.0714859
age	.0938182	.0237794	3.95	0.000	.0472115	.1404249
agesq	-.0010508	.0002735	-3.84	0.000	-.0015868	-.0005148
_cons	7.780893	.2291541	33.95	0.000	7.33176	8.230027

Instrumented: cigs
Instruments: educ age agesq lcigpric restaurn

Question 3.5

Table 7

```
. reg lincome cigs educ age agesq
```

Source	SS	df	MS
Model	67.5412888	4	16.8853222
Residual	341.854549	802	.426252555
Total	409.395838	806	.507935283

Number of obs =	807
F(4, 802) =	39.61
Prob > F =	0.0000
R-squared =	0.1650
Adj R-squared =	0.1608
Root MSE =	.65288

lincome	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
cigs	.0017306	.0017137	1.01	0.313	-.0016333	.0050945
educ	.0603606	.0078983	7.64	0.000	.0448567	.0758645
age	.0576908	.0076436	7.55	0.000	.042687	.0726946
agesq	-.0006306	.0000834	-7.56	0.000	-.0007943	-.0004669
_cons	7.795444	.1704271	45.74	0.000	7.460908	8.129979

Table 8

	(1)	(2)	(3)
	lincome	lincome	lincome
	b/se	b/se	b/se
cigs	-0.0421	-0.0414	0.0017
	(0.02614)	(0.02595)	(0.00171)
educ	0.0397**	0.0400**	0.0604***
	(0.01623)	(0.01611)	(0.00790)
age	0.0938***	0.0932***	0.0577***
	(0.02378)	(0.02361)	(0.00764)
agesq	-0.0011***	-0.0010***	-0.0006***
	(0.00027)	(0.00027)	(0.00008)
_cons	7.7809***	7.7811***	7.7954***
	(0.22915)	(0.22742)	(0.17043)
r2	.	.	0.165
N	807	807	807

* p<0.1, ** p<0.05, *** p<0.01