Assignment 5 - Yanyu Chen

1. Effect of Personal Computer

i.

Because PC ownership may be correlated with family income. A student from a richer family may be more likely to have a PC. Also, students from richer family may have better GPA because they more have additional educational resources provided by their family.

ii.

PC ownership is likely to be related to parents' annual income because they may have more spare money for their children to buy their own PC. In contrast, a PC may not be affordable for parents with lower annual income. Even if every student has a PC in that university, parents with higher income may provide better condition PC for their children.

It doesn't mean parental income is a good IV for PC. Like what explained in i, parents' income will not affect students' GPA only through the 'PC pathway', it can also affect students' GPA by providing other extra educational resources for them such as more books and private mentor.

iii.

The IV must satisfy three requirements:

In this case, these grants will probably affect the ownership of PC because it makes it more affordable.

It also satisfy the independence assumption, because it's randomly assigned to one-half of the students.

Also, these grants are probably affecting students' GPA only by affecting the ownership of PC. That's because receiving these grants are not likely to affect their GPA directly since it's only provided for buying PC, and the 'joy of receiving the grants' is not likely to affect their GPA in the long run too.

So, the IV can be 'having received the grants or not', if yes, the value is 1, else it's 0. And we can use 2SLS to calculate the LATE.

2. School Choice Program

i.

Because there may still exist other omitted variables. For example, the choice may be correlated with the students' ability and ambition. Even if this ability and ambition may also be correlated with faminc, it's unlikely they have a full correlation (corr=1), so the choice will still be correlated with u_1.

ii.

It's uncorrelated with u_1. Because there is already a variable named faminc, meaning that the faminc is already controlled from the u_1. So, within these controlled groups, the randomly assigned grants will satisfy the independent assumption because of the randomization. Also, it satisfies other two requirements by affecting students' scores only through affecting students' decision of attending a choice school.

iii.

Reduced form:

$$choice_i = \theta_0 + \theta_1 faminc_i + \theta_2 grant_i + \epsilon_i$$

To make sure grant is partially correlated with choice, the θ_2 can not be zero. And it's better that the t-statistic for coefficient θ_2 in the t-test is greater than 2 (or 1.96). Otherwise it will become a 'weak' IV.

iv.

Reduced form:

$$score_i = \beta_0 + \beta_1 faminc_i + \beta_2 grant_i + \gamma_i$$

In this equation, we can estimated the average effect of the grants without considering the 'choice'. Holding everything else constant, on average, 1 unit change of grant is associated with β_2 change of score. For example, holding family income constant, if the grant increases by 1, the score is estimated to increase by β_2 .

3.

i.

```
. count if selectyrs == 0
  468
```

468 students were never awarded a voucher.

```
. count if selectyrs == 4
  108
```

108 students had a voucher available for four years.

```
. count if choiceyrs == 4
56
```

56 students attended a choice school for four years.

ii.

```
. reg choiceyrs selectyrs
```

Source	SS	df	MS	Number of obs	=	990
				F(1, 988)	=	3712.53
Model	1231.71919	1	1231.71919	Prob > F	=	0.0000
Residual	327.791925	988	.331773203	R-squared	=	0.7898
				Adj R-squared	=	0.7896
Total	1559.51111	989	1.57685653	Root MSE	=	.576

choiceyrs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
selectyrs	.7668317	.0125853	60.93	0.000	.7421347	.7915288
_cons	.0199189	.0246115	0.81	0.419	-.0283778	.0682157

$\widehat{choiceyrs} = 0.02 + 0.767 selectyrs$

Yes, this is what I would expect. The number of years attending a choice school should be positively associated with the vouchers that students got to attend a choice school. In fact, we R-squared is not low and the t-statistic is about 61, which is very large enough to make sure the significance.

iii.

```
. reg mnce choiceyrs
```

Source	SS	df	MS	Number of obs	=	990
				F(1, 988)	=	12.22
Model	5262.76064	1	5262.76064	Prob > F	=	0.0005
Residual	425568.372	988	430.737218	R-squared	=	0.0122
				Adj R-squared	=	0.0112
Total	430831.132	989	435.622985	Root MSE	=	20.754

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	-1.837014	.5255475	-3.50	0.000	-2.868332	-.8056969
_cons	46.2344	.8507057	54.35	0.000	44.56501	47.9038

$\widehat{mnce} = 46.23 - 1.84 choiceyrs$

I found that the math NCE test score seems to be significantly (t = -3.5) negatively associated with the number of years attending a choice school.

This means on average, holding everything else constant, each year in a choice school is associated with a 1.84 decrease of the math test score.

It's not actually what I would expect because it's telling us that attending a choice school has negative effect on students math score. However, I can understand this result because it's just a simple regression and there may be other omitted variables that we have not included.

```
. reg mnce choiceyrs black hispanic female
```

Source	SS	df	MS	Number of obs	=	990
				F(4, 985)	=	23.40
Model	37384.9579	4	9346.23948	Prob > F	=	0.0000
Residual	393446.174	985	399.43774	R-squared	=	0.0868
				Adj R-squared	=	0.0831
Total	430831.132	989	435.622985	Root MSE	=	19.986

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	-.5652475	.5307292	-1.07	0.287	-1.606737	.4762423
black	-16.01743	1.794448	-8.93	0.000	-19.53881	-12.49605
hispanic	-13.40287	2.316797	-5.79	0.000	-17.9493	-8.856449
female	1.352745	1.275761	1.06	0.289	-1.150777	3.856267
_cons	57.12192	1.656697	34.48	0.000	53.87086	60.37298

$$\widehat{mnce} = 57.12 - 0.57 choiceyrs - 16.02 black - 13.40 hispanic + 1.35 female$$

After adding black, hispanic, and female into the regression, the coefficient of choiceyrs is still a negative number but its absolute value is smaller and it becomes not significant because the t-statistic is -1.07 which is smaller than 2.

In this situation, the result is showing that attending a choice school has no effect or association with students' math test score.

iv.

$$mnce_i = \beta_0 + \beta_1 choiceyrs_i + \beta_2 black_i + \beta_3 hispanic_i + \beta_4 female_i + u_{1i}$$

Because there may still exist selection bias even if the race and gender is controlled. For example, people with lower math score may be more likely to attend a choice school because they think it will help them to improve their math score. That is to say, on average, people that choose to attend a choice school may have lower math score than those who do not attend before the treatment.

This will lead to an underestimated treatment effect. Maybe this is the reason it's showing that attending a choice school has no effect or association with students' math test score.

v.

```
. ivreg mnce (choiceyrs = selectyrs) black hispanic female

Instrumental variables (2SLS) regression
```

Source	SS	df	MS		
Model	37236.1583	4	9309.03959	Number of obs	= 990
Residual	393594.974	985	399.588806	F(4, 985)	= 23.15
				Prob > F	= 0.0000
				R-squared	= 0.0864
				Adj R-squared	= 0.0827
Total	430831.132	989	435.622985	Root MSE	= 19.99

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	-.2413189	.6052706	-0.40	0.690	-1.429087	.9464492
black	-16.31692	1.814816	-8.99	0.000	-19.87827	-12.75557
hispanic	-13.7754	2.341246	-5.88	0.000	-18.3698	-9.180999
female	1.319709	1.276347	1.03	0.301	-1.184962	3.824381
_cons	57.06804	1.657716	34.43	0.000	53.81498	60.3211

```
Instrumented:  choiceyrs
Instruments:   black hispanic female selectyrs
```

$$\widehat{mnce} = 57.07 - 0.24 choiceyrs - 16.32 black - 13.78 hispanic + 1.32 female$$

No, the IV does not produce a positive effect of attending a choice school. The value of the coefficient for choiceyrs is still negative and the t-statistic is also smaller, meaning that it's still a negative number with no statistical significance. So, from this result it seems that attending a choice school still has no effect or association with students' math test score.

About the other 3 variables, the coefficients of black and hispanic are all negative and the t-statistics are -8.99 and -5.88, suggesting that this coefficient is statistically significant. Maybe this is because attending a choice school has different effects for students of different races.

About female, the t-statistic is smaller than 2 so the coefficient is not statistically significant, and we can say that the treatment effect has no difference between genders.

vi.

OLS:

```
. reg mnce choiceyrs black hispanic female mnce90
```

Source	SS	df	MS	Number of obs	=	328
				F(5, 322)	=	47.34
Model	60812.6587	5	12162.5317	Prob > F	=	0.0000
Residual	82727.2407	322	256.916896	R-squared	=	0.4237
				Adj R-squared	=	0.4147
Total	143539.899	327	438.959937	Root MSE	=	16.029

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	.4105823	.7358659	0.56	0.577	-1.03713	1.858294
black	-8.305183	2.546084	-3.26	0.001	-13.31424	-3.296123
hispanic	-4.10498	3.362347	-1.22	0.223	-10.71992	2.509962
female	-.882847	1.775972	-0.50	0.619	-4.37682	2.611126
mnce90	.6203655	.0484023	12.82	0.000	.5251409	.7155902
_cons	22.1529	3.620352	6.12	0.000	15.03037	29.27543

$$\widehat{mnce} = 22.15 + 0.41 choiceyrs - 8.31 black - 4.10 hispanic - 0.88 female + 0.62 mnce90$$

IV:

```
. ivreg mnce (choiceyrs = selectyrs) black hispanic female mnce90

Instrumental variables (2SLS) regression
```

Source	SS	df	MS	Number of obs	=	328
				F(5, 322)	=	47.64
Model	59897.5427	5	11979.5085	Prob > F	=	0.0000
Residual	83642.3567	322	259.758872	R-squared	=	0.4173
				Adj R-squared	=	0.4082
Total	143539.899	327	438.959937	Root MSE	=	16.117

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	1.799385	.8601938	2.09	0.037	.1070748	3.491694
black	-9.067109	2.571415	-3.53	0.000	-14.12601	-4.008214
hispanic	-5.00373	3.392791	-1.47	0.141	-11.67857	1.671107
female	-1.020484	1.786296	-0.57	0.568	-4.53477	2.493801
mnce90	.6288128	.0487424	12.90	0.000	.5329191	.7247065
_cons	21.53886	3.645484	5.91	0.000	14.36689	28.71084

```
Instrumented:  choiceyrs
Instruments:   black hispanic female mnce90 selectyrs
```

$$\widehat{mnce} = 21.54 + 1.80 choiceyrs - 9.07 black - 5.00 hispanic - 1.02 female + 0.63 mnce90$$

With OLS, although the coefficient of choiceyrs become positive. The t-statistic (0.56) still suggests that it's not statistically significant.

With IV, the coefficient of choiceyrs become greater positive, and the t-statistic (2.09) now suggests that it's statistically significant. From this result, the causal effect of each year in a choice school is an increasing math score by 1.8. For example, holding everything else constant, on average, spending 1 more year attending a choice school is associated with 1.8 higher math score.

I think this is a large effect because this effect is kind of 'cumulative', meaning that if attending a choice school for 4 years, on average, the math score is estimated to increase by 1.8 times 4 which equals to 7.2. This is quite large.

vii.

Because the number of observations gets much smaller when including the mnce90 in the regression. Before that, the number was 990, but after that, it's only 328.

The problem is that we don't know if the missing observations are lost randomly or with some patterns that we do not know. So, the regression is only convincing for the students that have mnce90 in the dataset, and for other students or the whole population, we can not be very sure about this regression.

4.

i.

Because this is a demand function, the sign of α_1 should be negative. (As airfare increases, the demand for air ticket is likely to decrease.)

ii.

```
. reg lpassen lfare ldist ldistsq if year == 1997
```

Source	SS	df	MS	Number of obs	=	1,149
				F(3, 1145)	=	22.96
Model	46.7630701	3	15.58769	Prob > F	=	0.0000
Residual	777.465948	1,145	.679009561	R-squared	=	0.0567
				Adj R-squared	=	0.0543
Total	824.229018	1,148	.717969528	Root MSE	=	.82402

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	-.3911724	.0672488	-5.82	0.000	-.5231171	-.2592277
ldist	-1.570478	.6285174	-2.50	0.013	-2.803653	-.3373028
ldistsq	.115756	.0477163	2.43	0.015	.0221349	.2093771
_cons	13.22961	2.100152	6.30	0.000	9.109035	17.35019

$$\widehat{log(passen)} = 13.23 - 0.39log(fare) - 1.57log(dist) + 0.12[log(dist)]^2$$

The coefficient of log(fare) is -0.39 with a large enough t-statistic absolute value. So, the estimated price elasticity is about -0.39.

iii.

We must assume that the demand is only correlated with the price (airfare) because when concen is exogenous, it should not affect the demand in any way except through affecting the price (airfare)

iv.

```
. reg lfare concen ldist ldistsq if year == 1997
```

Source	SS	df	MS		
Model	99.8444059	3	33.2814686	Number of obs	= 1,149
Residual	145.115611	1,145	.126738525	F(3, 1145)	= 262.60
				Prob > F	= 0.0000
Total	244.960016	1,148	.213379805	R-squared	= 0.4076
				Adj R-squared	= 0.4060
				Root MSE	= .356

lfare	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
concen	.3950364	.0627179	6.30	0.000	.2719815	.5180914
ldist	-.9360734	.2718439	-3.44	0.001	-1.469441	-.4027054
ldistsq	.10807	.0206224	5.24	0.000	.0676079	.148532
_cons	6.190051	.8898786	6.96	0.000	4.444075	7.936026

$\widehat{log(fare)} = 6.19 + 0.40concen - 0.94log(dist) + 0.11[log(dist)]^2$

Yes, concen has a positive effect on log(fare). The coefficient is about 0.4 and it's statistically significant because the t-statistic is 6.3.

v.

```
. ivreg lpassen (lfare = concen) ldist ldistsq if year == 1997

Instrumental variables (2SLS) regression
```

Source	SS	df	MS		
Model	-45.2473059	3	-15.0824353	Number of obs	= 1,149
Residual	869.476324	1,145	.759367968	F(3, 1145)	= 13.48
				Prob > F	= 0.0000
Total	824.229018	1,148	.717969528	R-squared	= .
				Adj R-squared	= .
				Root MSE	= .87142

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	-1.173998	.3886211	-3.02	0.003	-1.936487	-.4115084
ldist	-2.175665	.7273403	-2.99	0.003	-3.602735	-.7485961
ldistsq	.1870288	.0612885	3.05	0.002	.0667785	.3072792
_cons	18.01375	3.222482	5.59	0.000	11.69112	24.33638

```
Instrumented:  lfare
Instruments:   ldist ldistsq concen
```

$$\widehat{log(passen)} = 18.01 - 1.174log(fare) - 2.18log(dist) + 0.19[log(dist)]^2$$

Now the estimated price elasticity is about -1.174. Its absolute value is greater than that in the OLS, meaning that customers are more sensitive to the price changes. The absolute value of its t-statistic is smaller than that in the OLS, but it's still showing that the estimated price elasticity is statistically significant because its absolute value is still greater than 2.

Statistics/Data Analysis

```
      name:  <unnamed>
       log:  D:\Heinz\Econometrics 1\HW\HW5_log.smcl
  log type:  smcl
 opened on:  30 Nov 2021, 22:21:09

1 . br

2 . count if selectyrs == 0
  468

3 . count if selectyrs == 4
  108

4 . count if choiceyrs == 4
  56

5 . reg choiceyrs selectyrs

      Source |       SS           df       MS      Number of obs   =       990
-------------+----------------------------------   F(1, 988)       =   3712.53
       Model |  1231.71919         1  1231.71919   Prob > F        =    0.0000
    Residual |  327.791925       988  .331773203   R-squared       =    0.7898
-------------+----------------------------------   Adj R-squared   =    0.7896
       Total |  1559.51111       989  1.57685653   Root MSE        =      .576

------------------------------------------------------------------------------
   choiceyrs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   selectyrs |   .7668317   .0125853    60.93   0.000     .7421347    .7915288
       _cons |   .0199189   .0246115     0.81   0.419    -.0283778    .0682157
------------------------------------------------------------------------------

6 . reg mnce choiceyrs

      Source |       SS           df       MS      Number of obs   =       990
-------------+----------------------------------   F(1, 988)       =     12.22
       Model |  5262.76064         1  5262.76064   Prob > F        =    0.0005
    Residual |  425568.372       988  430.737218   R-squared       =    0.0122
-------------+----------------------------------   Adj R-squared   =    0.0112
       Total |  430831.132       989  435.622985   Root MSE        =    20.754

------------------------------------------------------------------------------
        mnce |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   choiceyrs |  -1.837014   .5255475    -3.50   0.000    -2.868332   -.8056969
       _cons |    46.2344   .8507057    54.35   0.000     44.56501    47.9038
------------------------------------------------------------------------------

7 . br

8 . reg mnce choiceyrs black hispanic female

      Source |       SS           df       MS      Number of obs   =       990
-------------+----------------------------------   F(4, 985)       =     23.40
       Model |  37384.9579         4  9346.23948   Prob > F        =    0.0000
    Residual |  393446.174       985   399.43774   R-squared       =    0.0868
-------------+----------------------------------   Adj R-squared   =    0.0831
       Total |  430831.132       989  435.622985   Root MSE        =    19.986

------------------------------------------------------------------------------
        mnce |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   choiceyrs |  -.5652475   .5307292    -1.07   0.287    -1.606737    .4762423
       black |  -16.01743   1.794448    -8.93   0.000    -19.53881   -12.49605
    hispanic |  -13.40287   2.316797    -5.79   0.000     -17.9493   -8.856449
      female |   1.352745   1.275761     1.06   0.289    -1.150777    3.856267
       _cons |   57.12192   1.656697    34.48   0.000     53.87086    60.37298
------------------------------------------------------------------------------
```

```
9 . ivreg mnce (choiceyrs = selectyrs) black hispanic female

Instrumental variables (2SLS) regression
```

Source	SS	df	MS		
				Number of obs	= 990
				F(4, 985)	= 23.15
Model	37236.1583	4	9309.03959	Prob > F	= 0.0000
Residual	393594.974	985	399.588806	R-squared	= 0.0864
				Adj R-squared	= 0.0827
Total	430831.132	989	435.622985	Root MSE	= 19.99

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	-.2413189	.6052706	-0.40	0.690	-1.429087	.9464492
black	-16.31692	1.814816	-8.99	0.000	-19.87827	-12.75557
hispanic	-13.7754	2.341246	-5.88	0.000	-18.3698	-9.180999
female	1.319709	1.276347	1.03	0.301	-1.184962	3.824381
_cons	57.06804	1.657716	34.43	0.000	53.81498	60.3211

```
Instrumented:  choiceyrs
Instruments:   black hispanic female selectyrs
```

```
10. reg mnce choiceyrs black hispanic female mnce90
```

Source	SS	df	MS		
				Number of obs	= 328
				F(5, 322)	= 47.34
Model	60812.6587	5	12162.5317	Prob > F	= 0.0000
Residual	82727.2407	322	256.916896	R-squared	= 0.4237
				Adj R-squared	= 0.4147
Total	143539.899	327	438.959937	Root MSE	= 16.029

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	.4105823	.7358659	0.56	0.577	-1.03713	1.858294
black	-8.305183	2.546084	-3.26	0.001	-13.31424	-3.296123
hispanic	-4.10498	3.362347	-1.22	0.223	-10.71992	2.509962
female	-.882847	1.775972	-0.50	0.619	-4.37682	2.611126
mnce90	.6203655	.0484023	12.82	0.000	.5251409	.7155902
_cons	22.1529	3.620352	6.12	0.000	15.03037	29.27543

```
11. ivreg mnce (choiceyrs = selectyrs) black hispanic female mnce90

Instrumental variables (2SLS) regression
```

Source	SS	df	MS		
				Number of obs	= 328
				F(5, 322)	= 47.64
Model	59897.5427	5	11979.5085	Prob > F	= 0.0000
Residual	83642.3567	322	259.758872	R-squared	= 0.4173
				Adj R-squared	= 0.4082
Total	143539.899	327	438.959937	Root MSE	= 16.117

mnce	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
choiceyrs	1.799385	.8601938	2.09	0.037	.1070748	3.491694
black	-9.067109	2.571415	-3.53	0.000	-14.12601	-4.008214
hispanic	-5.00373	3.392791	-1.47	0.141	-11.67857	1.671107
female	-1.020484	1.786296	-0.57	0.568	-4.53477	2.493801
mnce90	.6288128	.0487424	12.90	0.000	.5329191	.7247065
_cons	21.53886	3.645484	5.91	0.000	14.36689	28.71084

```
Instrumented:  choiceyrs
Instruments:   black hispanic female mnce90 selectyrs
```

```
12. avg(mnce)
  command avg is unrecognized
  r(199);

13. mean(mnce)

Mean estimation                   Number of obs    =       990

------------------------------------------------------------------------------
             |       Mean   Std. Err.      [95% Conf. Interval]
-------------+----------------------------------------------------------------
        mnce |   44.35657   .6633425       43.05485    45.65829
------------------------------------------------------------------------------

14. use "D:\Heinz\Econometrics 1\HW\HW5\AIRFARE.dta"

15. reg lpassen lfare ldist ldistsq

      Source |       SS           df       MS      Number of obs   =     4,596
-------------+----------------------------------   F(3, 4592)      =     99.01
       Model |  218.155759         3  72.7185863   Prob > F        =    0.0000
    Residual |  3372.53166     4,592  .734436336   R-squared       =    0.0608
-------------+----------------------------------   Adj R-squared   =    0.0601
       Total |  3590.68741     4,595  .781433605   Root MSE        =    .85699

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.5496928   .036824   -14.93   0.000    -.6218855   -.4775001
       ldist |  -1.537574   .3269891   -4.70   0.000    -2.17863    -.8965181
     ldistsq |   .1213552   .0248287    4.89   0.000     .072679     .1700315
       _cons |   13.61972   1.095649    12.43   0.000    11.47172    15.76772
------------------------------------------------------------------------------

16. reg lpassen lfare ldist ldistsq, robust

Linear regression                               Number of obs     =     4,596
                                                F(3, 4592)        =     97.63
                                                Prob > F          =    0.0000
                                                R-squared         =    0.0608
                                                Root MSE          =    .85699

------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.5496928   .0365518   -15.04   0.000    -.6213518   -.4780338
       ldist |  -1.537574    .352982    -4.36   0.000    -2.229588   -.8455597
     ldistsq |   .1213552   .0265173     4.58   0.000     .0693685    .1733419
       _cons |   13.61972   1.173176    11.61   0.000     11.31973    15.91971
------------------------------------------------------------------------------

17. reg lfare concern ldist ldistsq
  variable concern not found
  r(111);

18. reg lfare concen ldist ldistsq

      Source |       SS           df       MS      Number of obs   =     4,596
-------------+----------------------------------   F(3, 4592)      =   1015.85
       Model |  349.089591         3  116.363197   Prob > F        =    0.0000
    Residual |  526.004782     4,592  .11454808   R-squared       =    0.3989
-------------+----------------------------------   Adj R-squared   =    0.3985
       Total |  875.094374     4,595  .190444913   Root MSE        =    .33845
```

lfare	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
concen	.3526892	.0302101	11.67	0.000	.2934628	.4119157
ldist	-.899166	.1290132	-6.97	0.000	-1.152094	-.6462382
ldistsq	.1027463	.0097816	10.50	0.000	.0835697	.1219228
_cons	6.249577	.4229656	14.78	0.000	5.420361	7.078793

19. reg lfare concen ldist ldistsq, robust

Linear regression

Number of obs	=	4,596
F(3, 4592)	=	1095.70
Prob > F	=	0.0000
R-squared	=	0.3989
Root MSE	=	.33845

lfare	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
concen	.3526892	.0319869	11.03	0.000	.2899795	.415399
ldist	-.899166	.1418311	-6.34	0.000	-1.177223	-.6211089
ldistsq	.1027463	.0105257	9.76	0.000	.0821109	.1233817
_cons	6.249577	.4751374	13.15	0.000	5.31808	7.181075

20. br

21. reg lpassen lfare ldist ldistsq if year == 1997

Source	SS	df	MS		
Model	46.7630701	3	15.58769	Number of obs	= 1,149
Residual	777.465948	1,145	.679009561	F(3, 1145)	= 22.96
Total	824.229018	1,148	.717969528	Prob > F	= 0.0000
				R-squared	= 0.0567
				Adj R-squared	= 0.0543
				Root MSE	= .82402

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	-.3911724	.0672488	-5.82	0.000	-.5231171	-.2592277
ldist	-1.570478	.6285174	-2.50	0.013	-2.803653	-.3373028
ldistsq	.115756	.0477163	2.43	0.015	.0221349	.2093771
_cons	13.22961	2.100152	6.30	0.000	9.109035	17.35019

22. reg lpassen lfare ldist ldistsq concen if year == 1997

Source	SS	df	MS		
Model	49.9510984	4	12.4877746	Number of obs	= 1,149
Residual	774.277919	1,144	.676816363	F(4, 1144)	= 18.45
Total	824.229018	1,148	.717969528	Prob > F	= 0.0000
				R-squared	= 0.0606
				Adj R-squared	= 0.0573
				Root MSE	= .82269

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	-.3640486	.0682933	-5.33	0.000	-.4980429	-.2300544
ldist	-1.417494	.6314482	-2.24	0.025	-2.65642	-.178567
ldistsq	.0994977	.0482245	2.06	0.039	.0048792	.1941161
concen	-.3199594	.1474244	-2.17	0.030	-.6092119	-.030707
_cons	13.00012	2.099422	6.19	0.000	8.880973	17.11927

23. reg lfare concen ldist ldistsq if year == 1997

Source	SS	df	MS		
Model	**99.8444059**	**3**	**33.2814686**	Number of obs	= **1,149**
Residual	**145.115611**	**1,145**	**.126738525**	F(3, 1145)	= **262.60**
				Prob > F	= **0.0000**
Total	**244.960016**	**1,148**	**.213379805**	R-squared	= **0.4076**
				Adj R-squared	= **0.4060**
				Root MSE	= **.356**

lfare	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
concen	**.3950364**	**.0627179**	**6.30**	**0.000**	**.2719815**	**.5180914**
ldist	**-.9360734**	**.2718439**	**-3.44**	**0.001**	**-1.469441**	**-.4027054**
ldistsq	**.10807**	**.0206224**	**5.24**	**0.000**	**.0676079**	**.148532**
_cons	**6.190051**	**.8898786**	**6.96**	**0.000**	**4.444075**	**7.936026**

24. ivreg lpassen (lfare = concen) ldist ldistsq if year == 1997

Instrumental variables (2SLS) regression

Source	SS	df	MS		
Model	**-45.2473059**	**3**	**-15.0824353**	Number of obs	= **1,149**
Residual	**869.476324**	**1,145**	**.759367968**	F(3, 1145)	= **13.48**
				Prob > F	= **0.0000**
Total	**824.229018**	**1,148**	**.717969528**	R-squared	= **.**
				Adj R-squared	= **.**
				Root MSE	= **.87142**

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	**-1.173998**	**.3886211**	**-3.02**	**0.003**	**-1.936487**	**-.4115084**
ldist	**-2.175665**	**.7273403**	**-2.99**	**0.003**	**-3.602735**	**-.7485961**
ldistsq	**.1870288**	**.0612885**	**3.05**	**0.002**	**.0667785**	**.3072792**
_cons	**18.01375**	**3.222482**	**5.59**	**0.000**	**11.69112**	**24.33638**

Instrumented: lfare
Instruments: ldist ldistsq concen

25. log close
name: **<unnamed>**
log: **D:\Heinz\Econometrics 1\HW\HW5_log.smcl**
log type: **smcl**
closed on: **1 Dec 2021, 00:57:02**