Ulaş Eraslan 25058

ECON 301-PS5

1.1

```
. reg lpassen lfare ldistsq y00 y99 y98 ldist

      Source |       SS           df       MS      Number of obs   =     4,596
-------------+----------------------------------   F(6, 4589)      =     52.48
       Model |  230.557732         6  38.4262887   Prob > F        =    0.0000
    Residual |  3360.12968     4,589  .732213921   R-squared       =    0.0642
-------------+----------------------------------   Adj R-squared   =    0.0630
       Total |  3590.68741     4,595  .781433605   Root MSE        =    .85569

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.5647711   .0369644   -15.28   0.000    -.6372392   -.4923031
     ldistsq |   .1227088   .0247935     4.95   0.000     .0741017     .171316
         y00 |   .1380369   .0358761     3.85   0.000     .0677024    .2083713
         y99 |    .081651    .035724     2.29   0.022     .0116148    .1516873
         y98 |   .0321212   .0357118     0.90   0.368    -.0378911    .1021335
       ldist |   -1.54939   .3265076    -4.75   0.000    -2.189502   -.9092778
       _cons |   13.65144   1.094166    12.48   0.000     11.50635    15.79653
------------------------------------------------------------------------------
```

According to results, the effect of the increase in prices on demand is negative and since the t-value is smaller than -1.96 it is significant. The result can be interpreted as, 1% increase in average price of one-way ticket, decreases the average number of passengers per day by 0.56 %. So it can be said that since the ratio of percentage change in demand / percentage change in price is smaller than 1 (-0.56/1) the price elasticity of demand is inelastic.

In terms of years, in 1998 demand is 3.2% higher as compared to 1997. But result is not statistically significant.
In 1999, the demand is 8.1% higher as compared to 1997. Since t-value is 2.29 (>1.96) result is statistically significant.
In 2000, the demand is 13.8 % higher as compared to 1997. Since t-value is 3.85, result is statistically significant.

1.2

By using Fix Effect method, we eliminate the time-invariant unobserved individual characteristics that might be related with independent variables.
In this model, the price elasticity on demand can be interpreted as, 1% increase in average price of one-way ticket, decreases the average number of passengers per day by 1.15 %. So the price elasticity of demand is 1.15/1 which means that it is elastic. Since the p-value is around 0 the result is statistically significant. The result is higher as compared to the previous model which does not use Fixed Effect. Positive side of

```
R-sq:                                           Obs per group:
     within  = 0.4507                                         min =          4
     between = 0.0487                                         avg =        4.0
     overall = 0.0574                                         max =          4

                                                F(4,3443)         =     706.35
corr(u_i, Xb)  = -0.3249                        Prob > F          =     0.0000

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.155039   .0227645   -50.74   0.000    -1.199672   -1.110406
     ldistsq |          0  (omitted)
             |
        year |
       1998  |   .0464889   .0059898     7.76   0.000     .0347449    .0582329
       1999  |   .1023612   .0060174    17.01   0.000     .0905631    .1141592
       2000  |   .1946548   .0063513    30.65   0.000     .1822021    .2071075
             |
       ldist |          0  (omitted)
       _cons |   11.81677   .1151921   102.58   0.000     11.59092    12.04262
-------------+----------------------------------------------------------------
     sigma_u |  .89829067
     sigma_e |  .14295339
         rho |   .9753002   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(1148, 3443) = 141.20                Prob > F = 0.0000
```

using FE is it might eliminates the endogeneity issue by dropping the unobserved invariant variable which may correlated with explanatory variables. Therefore as it can be seen from the results, we can fix the underestimation in this case. So using Fixed Effect method is more consistent as compared to OLS estimation.

In FE estimation model omitted the distance and distance-squared variables because they are constant and not changing over-time so we eliminate their effects from model.

1.3 To consider “concen” as an instrumental variable the following conditions must be satisfied :

- It should not be directly related with the dependent variable in the model. In this context, “concen” should not be directly related with demand variable. The only channel that “concen” shows it’s effect is fare. There should be not any other channels.
- It should be highly correlated with endogenous variable. In this context, we may consider the ticket fares as an endogenous variable since it is affected by many other factors beside control variables.
- It should be uncorrelated with error term.

To check whether the “concen” variable satisfies the relevance assumption, we can use regression.

As it can be seen from the results, “concen“ is highly correlated with “lfare”. Since the t-value is 11.82 the result is statistically significant. Also we can say that lfare is the only channel that “concen” shows it affect on dependent variable. So we can use “concen” as an instrumental variable in our model.

```
. reg lfare concen y98 y99 y00 ldist ldistsq

      Source |       SS           df       MS      Number of obs   =     4,596
-------------+----------------------------------   F(6, 4589)      =    523.18
       Model |  355.453858         6  59.2423096   Prob > F        =    0.0000
    Residual |  519.640516     4,589  .113236112   R-squared       =    0.4062
-------------+----------------------------------   Adj R-squared   =    0.4054
       Total |  875.094374     4,595  .190444913   Root MSE        =    .33651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203   .0300691    11.98   0.000     .3011705    .4190702
         y98 |   .0211244   .0140419     1.50   0.133    -.0064046    .0486533
         y99 |   .0378496   .0140413     2.70   0.007      .010322    .0653772
         y00 |     .09987   .0140432     7.11   0.000     .0723385    .1274015
       ldist |  -.9016004    .128273    -7.03   0.000    -1.153077   -.6501235
     ldistsq |   .1030196   .0097255    10.59   0.000     .0839529    .1220863
       _cons |   6.209258   .4206247    14.76   0.000     5.384631    7.033884
------------------------------------------------------------------------------
```

1.4

I created a variable “lfarehat” which is based on the previous regression that utilized “concen” as an instrumental variable. So now, we eliminate most of the endogeneity problem.
Then, by using the new predicted "lfarehat” variable, I run the 2SLS estimation.

As it can be seen from the results,

```
. reg lpassen lfarehat y00 y99 y98 ldist ldistsq

      Source |       SS           df       MS      Number of obs   =     4,596
-------------+----------------------------------   F(6, 4589)      =     24.37
       Model |  110.891091         6  18.4818484   Prob > F        =    0.0000
    Residual |  3479.79632     4,589  .758290766   R-squared       =    0.0309
-------------+----------------------------------   Adj R-squared   =    0.0296
       Total |  3590.68741     4,595  .781433605   Root MSE        =     .8708

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lfarehat |  -1.776549   .2160716    -8.22   0.000    -2.200153   -1.352945
         y00 |   .2542695   .0418265     6.08   0.000     .1722695    .3362696
         y99 |   .1241675   .0371132     3.35   0.001     .0514079    .1969272
         y98 |   .0616171   .0367093     1.68   0.093    -.0103508    .1335851
       ldist |  -2.498972   .3717581    -6.72   0.000    -3.227797   -1.770147
     ldistsq |   .2314933   .0316459     7.32   0.000     .1694521    .2935344
       _cons |   21.21249   1.732746    12.24   0.000     17.81547     24.6095
------------------------------------------------------------------------------
```

the 1% increase in average price of one-way ticket, decreases the average number of passengers per day by 1.77 % with t-value -8.22 which is statistically significant. Elasticity is 1.77/1 since it is greater than 1 price elasticity of demand is elastic. It was inelastic in the simple regression model. This results can show the casual relationship between price and demand in a safer way as compared to previous models.

If we use “ivreg” command, the results are the following:

. ivreg lpassen ldistsq y98 y99 y00 ldist (lfare = concen)

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	-556.334915	6	-92.7224858
Residual	4147.02233	4,589	.903687586
Total	3590.68741	4,595	.781433605

Number of obs	=	4,596
F(6, 4589)	=	20.45
Prob > F	=	0.0000
R-squared	=	.
Adj R-squared	=	.
Root MSE	=	.95062

lpassen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lfare	-1.776549	.2358788	-7.53	0.000	-2.238985	-1.314113
ldistsq	.2314932	.0345468	6.70	0.000	.1637648	.2992216
y98	.0616171	.0400745	1.54	0.124	-.0169481	.1401824
y99	.1241675	.0405153	3.06	0.002	.044738	.2035971
y00	.2542695	.0456607	5.57	0.000	.1647525	.3437865
ldist	-2.498972	.4058371	-6.16	0.000	-3.294607	-1.703336
_cons	21.21249	1.891586	11.21	0.000	17.50407	24.9209

Instrumented: lfare
Instruments: ldistsq y98 y99 y00 ldist concen

The main difference between “ivreg” method and 2 step method is the standard error. We find the more accurate standard error as compare to 2 step method. And therefore the t-values are also different in this model. It might affect the significance levels as well, for instance change in y99 and y98. However these changes do not affect the importance of variables in this context.

2.1

. reg children educ age agesq

Source	SS	df	MS
Model	12243.0295	3	4081.00985
Residual	9284.14679	4,357	2.13085765
Total	21527.1763	4,360	4.93742577

Number of obs	=	4,361
F(3, 4357)	=	1915.20
Prob > F	=	0.0000
R-squared	=	0.5687
Adj R-squared	=	0.5684
Root MSE	=	1.4597

children	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.0905755	.0059207	-15.30	0.000	-.102183	-.0789679
age	.3324486	.0165495	20.09	0.000	.3000032	.364894
agesq	-.0026308	.0002726	-9.65	0.000	-.0031652	-.0020964
_cons	-4.138307	.2405942	-17.20	0.000	-4.609994	-3.66662

According to results, education level of a woman has a negative relationship with the number of children that she has. Since it’s t-value is smaller than -1.96 (p-value is smaller than 1% significance level) the effect of education on number of children is significant. Extra one year of education decreases the number of children that she has by 0.0905.
Age has a positive effect on education until a point (it is not infinite). According to results each extra increase in age, increases the number of children that she has by 0.33. Since t-value is 20.09 (p-value < 0.01) the effect is statistically significant.
As stated above, the effect of age is not infinite, to see that we used square of age. So the the effect of age is decreasing after a certain point (at age 83). The effect of extra one

year on age, decreases the number of children that a woman has by 0.002. Since it's t-value is smaller than -1.96 (p-value <0.01) this effect is statistically significant.

If 100 women receive another year of education, we expect 9 (9.05) children less among these women.

2.2 If being born in first-half affects the education level, we can use regression to to check the relationship between education and first-half born.

```
. reg educ frsthalf agesq age

      Source |       SS           df       MS      Number of obs   =     4,361
-------------+----------------------------------   F(3, 4357)      =    175.21
       Model |  7238.42472         3  2412.80824   Prob > F        =    0.0000
    Residual |   60001.141     4,357  13.7712052   R-squared       =    0.1077
-------------+----------------------------------   Adj R-squared   =    0.1070
       Total |  67239.5657     4,360  15.4219187   Root MSE        =     3.711

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    frsthalf |  -.8522854   .1128296    -7.55   0.000    -1.073489   -.6310821
       agesq |  -.0005056   .0006929    -0.73   0.466    -.0018641    .0008529
         age |  -.1079504   .0420402    -2.57   0.010    -.1903706   -.0255302
       _cons |   9.692864   .5980686    16.21   0.000     8.520346    10.86538
------------------------------------------------------------------------------
```

As it can be seen, the effect of first-half on education is statistically significant (p-value <0.01) and if a women born in the first-half of the year then she gets almost 1 year less education.
In addition to that, we should also investigate whether there are any other channel that is affected by being born in the first half of the year. I think the only channel that is affected by the first-half born is the education. Therefore, when we use the education as independent variable in children formula, we can use first-half as an instrumental variable to prevent from endogeneity.

2.3

```
. predict educhat
(option xb assumed; fitted values)

. reg children educhat age agesq

      Source |       SS           df       MS      Number of obs   =     4,361
-------------+----------------------------------   F(3, 4357)      =   1751.10
       Model |    11767.45         3  3922.48332   Prob > F        =    0.0000
    Residual |  9759.72638     4,357  2.24001064   R-squared       =    0.5466
-------------+----------------------------------   Adj R-squared   =    0.5463
       Total |  21527.1763     4,360  4.93742577   Root MSE        =    1.4967

------------------------------------------------------------------------------
    children |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educhat |   -.171499   .0533921    -3.21   0.001    -.2761746   -.0668233
         age |   .3236052    .017931    18.05   0.000     .2884514     .358759
       agesq |  -.0026723   .0002808    -9.52   0.000    -.0032228   -.0021218
       _cons |  -3.387805   .5503399    -6.16   0.000    -4.466751   -2.308858
------------------------------------------------------------------------------
```

To use the first-half born as an instrumental variable, I regressed education on first-half variable and used year and year-squared as control variables. The result that produced is called as “educthat”. Since this variable is only determined by first-half born variable, the endogeneity problem is purified.

The results of the model that I used educhat shows that the effect of education on number of children is again negative but effect is higher as compared to previous model that I used educ variable without instrumental variable (previous one is underestimated). According to new results, extra one year of eduction of women in Botswana , decreases the number of children by 0.17 (for 100 women educated, we expect 17 children less.) Since the t-value is smaller than -1.96, effect is statistically significant (p-value < 0.05).

2.4 When I add the binary variables, the coefficient of the educ in OLS is around -0.07 (negative effect on number of children) with t-value -11.82 which is significant. So it means that extra one year of education, decreases the number of children by 0.07 (for 100 women it is around 7 less children).
On the other hand, in 2SLS the coefficient of education is around -0.17 (negative effect on number of children) with t-value -2.42 which is significant at 5% significance level. It can be interpreted like extra one year of education on a women in Botswana, decreases the number of children 0.17. (For 100 women, it means 17 less children).
As it can be seen from the results, when we use instrumental variable we eliminate the endogeneity problem in the model. In addition, the underestimated effect of education on number of children is cleared as well.

When we consider the effect of having TV, in OLS model, tv coefficient is around -0.25 with t-value -2.78 which is significant. It means that if a woman has a TV, the number of child that she has is 0.25 less as compared the one who has not a TV (For 100 women who has TV, they have 25 less children as compared to ones who have not TV). In 2SLS the coefficient of TV is -0.017 with t-value —0.09 which is statistically insignificant. Hence it means that TV has not an important effect on number of children.
The relationship between TV ownership and having children can caused by the time allocation. For instance, parents may spend their spare time on TV more. Or they can get some more information about birth control via TV ads and some public informative channels.