Solutions to Advanced Econometrics Homework 2

Ashwin Iyenggar (1521001)

ashwin.iyenggar15@iimb.ernet.in

October 7, 2016

Question I (a)

I a(1): Recall the wage-setting equation from the education model discussed in class: ln wi = ai+ bSi. Interpret the parameter b in this equation.

Since we are regressing log wages on years of education, this means that $\%\Delta(wi) = (100 * b) * \Delta(Si)$. Therefore the economic meaning of the slope coefficient b is that a unit change in number of years of education, leads to a 100*b% increase in wages. That would have ideally been the case had Si been an exogenous regressor of ln wi. We discuss why that may not quite be the case in the following question. Table 2 nevertheless presents the regression results from this regression under model (1) where yeduc shows a positive and significant effect on the logarithm of hourly wages

```
reg lhwage yeduc
outreg2 using 'imagepath'Ia.tex, ctitle(baseline) tex(pretty frag) dec(4)
    replace
```

I a(2): Why might the coefficient on an OLS regression of log wages on years of schooling not be a good estimate of b?

The first question we ask ourselves is if the model presented in the previous question is well specified. First, we note that since the model does not account for ability, it is therefore quite possible that the error term in the above specification could also embody the effects of ability. Since ability may reasonably be expected to directly affect the number of years of schooling (assuming that those with higher ability may also tend to have higher years of schooling), we now have the additional problem of endogenous regressor in number of years of schooling. Clearly, the issues of model misspecification and omitted variables both lead to a poor estimate for b in the previously specified wage setting equation. Identifying a good instrumental variable that is correlated with the regressor (number of years of schooling), but that is not believed to be directly correlated with log wages (except for the effect through the number of years of schooling) may help with isolating some variation in number of years of schooling that is unrelated to the log of wages.

Question I (b)

I b: 2SLS and Instrument Variables. State clearly the requirements for these to be good instruments? Explain briefly why you might think that these variables satisfy those two requirements.

Building on top of our discussion in the previous section, we identify the following requirements for a good instrumental variable. First, the instrumental variable should affect the endogenous regressor variable (in our case, the number of years of schooling) Second, the instrumental variable should itself not appear as a regressor in the structural equation. In other words, the instrumental variable should not be correlated with the error term of the structural equation. Third, the instrumental variable should demonstrate variance that may be taken to be as good as assigned randomly. We note that while we may test for the first condition through the correlation matrix, conditions two and three are not things that may be tested from the data. Instead these need to be borne out from theory and an understanding of phenomena.

The question suggests that `z62 - z72` may be used as instrumental variables. From Table 1, we note that with the exception of `z62, z65, z71` all other z variables are significantly correlated with `yeduc`. We therefore have support for our first requirement. As noted previously, the second and third requirements however are hard to prove with data. Here we choose to rely on our understanding of the phenomena. The YOB dummies `d62 - d72` may themselves not be expected to be strongly correlated with earnings but it may be possible that the years of birth capture the stage of the business cycle at which people were hired, and therefore in someway determines the wages earned. A similar argument may be made for the program intensity of the INPRES program, where the argument may be that the well performing regions (as captured by the program intensity in the region) may have themselves determined the future earnings of its students. However, when multiplied the resulting value of the variables `z62 - z72` are expected to be uncorrelated with earnings as they capture neither a business cycle effect nor a regional effect completely. Finally, the creation of the interaction term provides us with a wider range of values for the distribution of the instrumental variables, thereby strengthening our requirement 3, though not satisfying it completely. We therefore conclude that the variables `z62 - z72` serve as strong instruments for the number of years of schooling in the structural equation where ln wages is modeled as a function of the number of years of schooling [1].

```
forvalues year=62/72 {
        gen d'year'= (YOB == 'year')
        }
forvalues year=62/72 {
        gen z'year'= d'year' * prog_int
        }
estpost cor yeduc z*
esttab . using 'imagepath'Ib.tex, title(Correlation of years of education
    to interaction of YOB dummy and Program Intensity \label{Ib}) label
    replace
```

---

[1]Prof. Shailender Swaminathan had noted in class that researchers have in the past used the lack of correlation between the instrument and other independent and control variables as a possible alternative to demonstrate the strength of the instrumental variable in that it satisfies our condition 2. However, we find that our argument lends stronger support to the requirement than such a correlation table

Table 1: Correlation of years of education to interaction of YOB dummy and Program Intensity

	(1) years of education
z62	0.00717
z63	0.0146***
z64	0.0201***
z65	0.00142
z66	0.0352***
z67	0.0343***
z68	0.0227***
z69	0.00896*
z70	-0.0152***
z71	-0.00234
z72	-0.0117**
Observations	59938

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

I b(i): For control variables in our regressions, we will use number of children in 1971, year-of-birth dummies, and year-of-birth dummies interacted with number of children in 1971 ("interacted with" means "multiplied by")

Control Variables are `ch71`, YOB dummies and the interaction of the two (generated using the Stata xi command)

I b(ii): Calculate the 2SLS estimate of b

I b(ii)(1):　First, use the instruments and control variables to predict years of education (i.e., run an OLS regression of years of education on the instruments and control variables, and form the predicted values of education from this regression).

```
xi: reg yeduc z* i.YOB*ch71
predict yeduchat
```

I b(ii)(2): Regress log hourly wages on this predicted value of years of education and the control variables. The coefficient on the predicted years of education is the 2SLS estimate of b. Compute and interpret the 2SLS estimate of b.

```
xi: reg lhwage yeduchat i.YOB*ch71
outreg2 using 'imagepath'Ia.tex, ctitle(man2sls) drop (_*) tex(pretty
    frag) dec(4) append
xi: ivregress 2sls lhwage (yeduc = z*) i.YOB*ch71
outreg2 using 'imagepath'Ia.tex, title("Comparing Baseline, Manual 2SLS,
    ivreg 2sls Results") ctitle(2sls) drop (_*) tex(pretty frag) dec(4)
    append
estat endog
estat firststage
estat overid
```

Table 2 shows the results of the 2SLS preformed in two stages by us in model 2, and that performed by the Stata command `ivregress` in model 3. We note that the coefficient estimates on both are the same at 0.0962, implying that a unit change in number of years of schooling is associated with a 9.62% increase in hourly wages.

I b(ii)(3): You are asked to write a short report on the costs versus benefits of the school construction program. What information/data might you need to write this report?

Our model here captures the wage related benefits to additional years of schooling. However, in order to write a report on the costs versus the benefits of schooling, we will need to consider

Table 2

Comparing Baseline, Manual 2SLS, ivreg 2sls Results

VARIABLES	(1) baseline	(2) man2sls	(3) 2sls
yeduchat		0.0962***	
		(0.0113)	
ch71		-0.0000	-0.0000
		(0.0000)	(0.0000)
yeduc	0.0792***		0.0962***
	(0.0006)		(0.0099)
Constant	6.1803***	6.2561***	6.2561***
	(0.0064)	(0.1090)	(0.0957)
Observations	59,938	59,938	59,938
R^2	0.2155	0.0507	0.2679

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

a few other things. First, we need to consider the costs of the school construction program. We will need to understand at the level of the region, not only the program intensity but also the capital costs of running that program. It would additionally be useful to capture data on alternative uses of this capital and expected returns to society from such spend. Costs also are a function of time, but that might be making things much more complicated. Second, in our model above we have considered only the immediate value to education as perceived by wages earned. It may not be unimaginable to expect that a higher skilled workforce would have increasing returns to education through the creation of higher value jobs through innovation and entrepreneurship. Increased economic activity may eventually lead to higher tax collections, and that may offset the calculation of cost of executing the program.

Question II (a)

```
cd /Users/aiyenggar/OneDrive/code/econometrics
local imagepath /Users/aiyenggar/OneDrive/code/articles/adv-eco-hw2-
    images/
use pollution_labels, clear

local mainctrl dincome dunemp dmnfcg ddens dwhite dfeml dage65 dhs ///
        dcoll durban dpoverty blt1080 blt2080 bltold80 dplumb vacant70 ///
        vacant80 vacrnt70 downer drevenue dtaxprop depend deduc dwelfr ///
        dhlth dhghwy

local polyctrl white2 coll3 vacant2 taxprop3 pcthlth2 white3 unemp2 ///
        vacant3 epend2 pcthlth3 femal2 unemp3 owner2 epend3 built102 ///
        femal3 mnfcg2 owner3 pcteduc2 built103 age2 mnfcg3 plumb2 ///
        pcteduc3 built202 pop2 age3 income2 plumb3 pcthghw2 built203 ///
        pop3 hs2 income3 revenue2 pcthghw3 builold2 urban2 hs3 ///
        poverty2 revenue3 pctwelf2 builold3 urban3 coll2 poverty3 ///
        taxprop2 pctwelf3

local keepmain dgtsp dincome dunemp dwhite dpoverty
local keeppoly income3 builold2
```

II a(1): Estimate the relationship between changes in air pollution and housing prices not adjusting for any control variables

---

```
reg dlhouse dgtsp
outreg2 using 'imagepath'IIa.tex, ctitle(baseline) tex(pretty frag) dec
    (4) replace
```

II a(2): Estimate the relationship between changes in air pollution and housing prices adjusting for the main effects of the control variables listed above

---

```
reg dlhouse dgtsp 'mainctrl'
outreg2 using 'imagepath'IIa.tex, ctitle(mainctrl) keep('keepmain') tex(
    pretty frag) dec(4) append
```

II a(3): Estimate the relationship between changes in air pollution and housing prices adjusting for the main effects and polynomials of the control variables included in the data set

---

```
reg dlhouse dgtsp 'mainctrl' 'polyctrl'
```

Table 3

Simple Regression (models 1-3) vs 2SLS Instrumenting with tsp7576 (models 4-6)

VARIABLES	(1) baseline	(2) mainctrl	(3) mainpoly	(4) baseline	(5) mainctrl	(6) mainpoly
dgtsp	0.0010***	0.0002	0.0001	-0.0037**	-0.0022**	-0.0027***
	(0.0002)	(0.0002)	(0.0002)	(0.0014)	(0.0009)	(0.0010)
dincome		0.0001***	0.0001***		0.0001***	0.0001***
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
dunemp		-0.8137***	-1.7407***		-0.8244***	-1.9775***
		(0.2241)	(0.4120)		(0.2441)	(0.4617)
dwhite		-0.5759***	-0.4610***		-0.5547***	-0.3770**
		(0.0894)	(0.1353)		(0.0977)	(0.1520)
dpoverty		-0.6130***	-0.9110***		-0.6876***	-0.8305**
		(0.1604)	(0.3455)		(0.1769)	(0.3821)
downer		-0.0953	-0.5217***		-0.0946	-0.4811**
		(0.0895)	(0.1847)		(0.0975)	(0.2042)
owner2			-5.2890***			-4.8432***
			(1.5944)			(1.7654)
age2			51.6218***			52.4923***
			(13.3330)			(14.7103)
poverty2			-10.1501			-9.3690
			(6.4216)			(7.0886)
Constant	0.2822***	0.2949***	0.3303**	0.2453***	0.2295***	0.2771*
	(0.0057)	(0.0595)	(0.1439)	(0.0129)	(0.0691)	(0.1598)
Observations	1,000	995	995	1,000	995	995
R^2	0.0166	0.5534	0.6321		0.4548	0.5153

Standard errors in parentheses

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

```
outreg2 using 'imagepath'IIa.tex, ctitle(mainpoly) keep('keepmain' '
    keeppoly') tex(pretty frag) dec(4) append
```

II a(4): What do your estimates imply and do they make sense?

Table 3 shows the results of the regression for the three models above in models 1 - 3. Kindly note that Table 3 as all future tables do not show the coefficient estimates for each of the variables in the model, but only a few indicative variables have been indicated. That data is made available in the associated log file aihw2.log that has also been submitted with this document. For the baseline model in Table 3, we note a small but significant positive coefficient estimate on `dgtsp` in model 1 where no controls are exercised. However once either the main controls (model 2) or the main controls and polynomial controls are exercised, we note that the coefficient estimate on `dgtsp` in models 2 and 3 become insignificant. This is hardly surprising since model may suffer from a misspecification, endogeneity or omitted variable problem.

II a(5): Describe the potential omitted variables biases.

The differences in economic indicators included in our previous model does not capture the underlying industrial and housing trends that may be behind the economic trends. It is imaginable that a variable as the nature of dominant technology might be highly correlated with the error term of the structural equation, thus implying an omitted variable bias. In this case the coefficient estimates on `dgtsp` may not be reliable.

II a(6): Provide indirect evidence on this ? i.e., the association between at least observable measures of economic shocks (dincome, dunemp, dmnfcg, ddens, blt1080) and changes in air pollution.

```
estpost cor dgtsp dincome dunemp dmnfcg ddens blt1080
esttab . using 'imagepath'IIa6.tex, title(Correlation of changes in
    pollution level to changes in economic indicators) replace
```

Table 4 depicts the correlations between the observable measures of economic shocks and the changes in air pollution. We note here that our main regressor `dgtsp` is highly correlated with several of the economic indicators. This is a clear indication of the possibility of their being another variable, not included in our model that may be affecting the levels of air pollution as well as the economic indicators.

Question II (b)

II b(1): To address the concern in part a), you might want to use federal EPA pollution regulation as a potential instrumental variable for pollution changes during the 1970s.

We will use `tsp7576` as our instrumental variable for reasons depicted in the following questions.

Table 4: Correlation of changes in pollution level to changes in economic indicators

| | (1) |
	dgtsp
dincome	0.107***
dunemp	-0.0913**
dmnfcg	0.149***
ddens	0.0697*
blt1080	0.124***
N	1000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

II b(2): What are the assumptions required for 1975-1976 regulatory status (tsp7576) to be a valid instrument for pollution changes when the outcome of interest is housing price changes?

We primarily need the instrument `tsp7576` to affect the endogenous regressor `dgtsp` but not itself appear as an exogenous regressor in the structural equation where the dependent variable is `dlhouse`. As depicted in Table 5 we notice that we have support for the first condition where the correlation is negative and significant. On the other hand, we notice a slightly positive and significant correlation with the dependent variable also. However, we continue with the use of `tsp7576` as the instrumental variable on account of the exogenous nature of the regulation and the impact of this on counties that were on the treshold.

```
estpost cor tsp7576 dlhouse dgtsp
esttab . using 'imagepath'IIb2.tex, title(Correlation of 1975-1976
    regulatory status with changes in housing prices and pollution levels\
    label{IIb2}) replace
```

II b(3): Provide indirect evidence on the validity of instrument ? e.g., relationship between the observable measures of economic shocks and the regulatory status (tsp7576).

Table 6 refers to the correlation of the instrumental variable `tsp7576` with observable measures of economic shocks and the regulator status. Clearly, we find that our instrumental variable is uncorrelated with any of these indicators, thereby strengthening its claim as an instrument. Given this, we argue that `tsp7576` may be unlikely to be significantly correlated with the error

Table 5: Correlation of 1975-1976 regulatory status with changes in housing prices and pollution levels

	(1) tsp7576
dlhouse	0.0947**
dgtsp	-0.198***
N	1000

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

term of the structural equation.

```
estpost cor tsp7576 dincome dunemp dmnfcg ddens blt1080
esttab . using 'imagepath'IIb3.tex, title(Correlation of regulatory
    status to changes in economic indicators\label{IIb3}) replace
```

Table 6: Correlation of regulatory status to changes in economic indicators

	(1) tsp7576
dincome	0.0334
dunemp	0.00680
dmnfcg	-0.00530
ddens	-0.0220
blt1080	-0.0296
N	1000

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

II b(4): Interpret your findings.

Table 5 and Table 6 indicate that our instrumental variable `tsp7576` is significantly correlated with the endogenous variable `dgtsp` but uncorrelated with other economic indicators. We may therefore conclude that the instrumental variable `tsp7576` is not significantly correlated with

the error term of the structural equation, and may therefore be used as an instrumental variable in this case.

Question II (c)

II c(1): Report the "first-stage" regression of air pollution changes on regulatory status (tsp7576) and the "reduced-form" regression of housing price changes on regulatory status (tsp7576), using the same three specifications you used in part a).

```
reg dgtsp tsp7576
outreg2 using 'imagepath'IIc1f.tex, ctitle(baseline) tex(pretty frag) dec
    (4) replace
reg dlhouse tsp7576
outreg2 using 'imagepath'IIc1r.tex, ctitle(baseline) tex(pretty frag) dec
    (4) replace

reg dgtsp tsp7576 'mainctrl'
outreg2 using 'imagepath'IIc1f.tex, ctitle(mainctrl) keep('keepmain') tex
    (pretty frag) dec(4) append
reg dlhouse tsp7576 'mainctrl'
outreg2 using 'imagepath'IIc1r.tex, ctitle(mainctrl) keep('keepmain') tex
    (pretty frag) dec(4) append

reg dgtsp tsp7576 'mainctrl' 'polyctrl'
outreg2 using 'imagepath'IIc1f.tex, title("First Stage Regression with
    Instrument as tsp7576") ctitle(mainpoly) keep('keepmain' 'keeppoly')
    tex(pretty frag) dec(4) append
reg dlhouse tsp7576 'mainctrl' 'polyctrl'
outreg2 using 'imagepath'IIc1r.tex, title("Reduced Form Regression with
    Instrument as tsp7576") ctitle(mainpoly) keep('keepmain' 'keeppoly')
    tex(pretty frag) dec(4) append
```

Table 7 displays the selected results of the first stage regression for each of the three model specifications suggested. Table 8 displays the selected results of the reduced form regression of dlhouse on tsp7576 for each of the three model specifications suggested.

II c(2): Interpret your findings

In the first stage regression as shown in Table 7, we observe that the effect of the instrumental variable tsp7576 on dgtsp is negative and significant in all the three models. This implies that that the regulation had a negative and significant effect on the pollution levels. We also observe that in models 1 and 2, other economic indicators are not statistically significant. This strengthens the case for tsp7576 as an instrumental variable in that most of the variance in dgtsp is captured by the instrumental variable.

The results from the reduced form regressions are demonstrated in Table 8

Table 7

First Stage Regression dgtsp on Instrument tsp7576

VARIABLES	(1) baseline	(2) mainctrl	(3) mainpoly
tsp7576	-9.8540***	-10.2397***	-9.4187***
	(1.5441)	(1.6195)	(1.6538)
dincome		0.0023	0.0047
		(0.0015)	(0.0050)
dunemp		15.2784	-67.4137
		(41.7999)	(81.0219)
dwhite		-9.6793	15.4823
		(16.8839)	(26.6990)
dpoverty		-3.8557	53.5235
		(30.1160)	(68.0748)
downer		9.4749	22.5692
		(16.7150)	(36.3480)
owner2			157.4626
			(313.4876)
age2			809.2212
			(2,623.2958)
poverty2			173.3820
			(1,262.8671)
Constant	-5.1013***	-21.1862*	-6.1976
	(0.8171)	(11.0641)	(28.3783)
Observations	1,000	995	995
R^2	0.0392	0.0983	0.1699

Standard errors in parentheses

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

Table 8

Reduced Form Regression dlhouse on Instrument tsp7576

VARIABLES	(1) baseline	(2) mainctrl	(3) mainpoly
tsp7576	0.0364***	0.0230***	0.0258***
	(0.0121)	(0.0087)	(0.0084)
dincome		0.0001***	0.0001**
		(0.0000)	(0.0000)
dunemp		-0.8587***	-1.7926***
		(0.2241)	(0.4099)
dwhite		-0.5330***	-0.4195***
		(0.0905)	(0.1351)
dpoverty		-0.6790***	-0.9772***
		(0.1615)	(0.3444)
downer		-0.1158	-0.5430***
		(0.0896)	(0.1839)
owner2			-5.2750***
			(1.5862)
age2			50.2730***
			(13.2732)
poverty2			-9.8445
			(6.3898)
Constant	0.2642***	0.2770***	0.2941**
	(0.0064)	(0.0593)	(0.1436)
Observations	1,000	995	995
R^2	0.0090	0.5558	0.6358

Standard errors in parentheses

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

II c(3): How does two-stage least squares use these two equations?

In the structural equation, when the endogenous regressor is replaced by its specification from the first stage equation, we get an equation where the regressand is expressed as a function of the instrumental variables and other variables. It turns out that the coefficient estimate in 2SLS for the effect of years of schooling on wages is nothing but the ratio of the coefficient estimate of the instrumental variable in the reduced form regression to the coefficient estimate of the instrumental variable in the first stage regression. We verify that this is indeed the case for each of the three model specifications as follows in Table 9.

	baseline	**mainctrl**	**mainpoly**
Reduced Form `tsp7576` coefficient estimate (α, Table 8)	0.0364	0.0230	0.0258
First Stage `tsp7576` coefficient estimate (β, Table 7)	-9.8540	-10.2397	-9.4187
α/β	-0.0037	-0.0022	-0.0027
2SLS coefficient estimate for `dgtsp` (Table 3)	-0.0037	-0.0022	-0.0027

Table 9: Calculating 2SLS Estimates from First Stage and Reduced Form Regressions

II c(4): Now estimate the effect of air pollution changes on housing price changes by two-stage least squares using the regulatory status indicator (tsp7576) as an instrument for the three specifications. Interpret the results.

We use the Stata command `ivregress 2sls` to perform the two stage least squares regression with instrumental variables for the sake of brevity. The results are presented in both Tables 3 and 10 to contrast the results of instrumenting with `tsp7675` against no instrument (Table 3) and against instrumenting with `tsp75` (Table 10)

```
ivregress 2sls dlhouse (dgtsp = tsp7576)
outreg2 using 'imagepath'IIa.tex, ctitle(baseline) tex(pretty frag) dec
    (4) append
outreg2 using 'imagepath'IIc.tex, ctitle(baseline) tex(pretty frag) dec
    (4) replace
ivregress 2sls dlhouse (dgtsp = tsp7576) 'mainctrl'
outreg2 using 'imagepath'IIa.tex, ctitle(mainctrl) keep('keepmain') tex(
    pretty frag) dec(4) append
outreg2 using 'imagepath'IIc.tex, ctitle(mainctrl) keep('keepmain') tex(
    pretty frag) dec(4) append
ivregress 2sls dlhouse (dgtsp = tsp7576) 'mainctrl' 'polyctrl'
outreg2 using 'imagepath'IIa.tex, title("Simple Regression (models 1-3)
    vs Instrumenting with tsp7576 (models 4-6)") ctitle(mainpoly) keep('
    keepmain' 'keeppoly') tex(pretty frag) dec(4) append
outreg2 using 'imagepath'IIc.tex, ctitle(mainpoly) keep('keepmain' '
    keeppoly') tex(pretty frag) dec(4) append
```

```
estat endog
estat firststage
```

We observe in Table 3 that models 4, 5 and 6 all show negative and significant coefficient estimates for `dgtsp`. This implies that housing prices are negatively correlated with change in pollution levels. This is indeed the result we would expect. However we also note that in models 1, 2, and 3 which did not use `tsp7576` as an instrumental variable, the coefficient estimates for `dgtsp` were positive for each of the three models but significant only for model 1. As we have argued before, this model suffered from endogeneity and an omitted variable bias. We are confident, however about the robustness of the estimates in models 4, 5 and 6.

II c(5): Now do the same analysis using the 1975 regulatory status indicator (tsp75) as an instrumental variable.

We again use the Stata command `ivregress 2sls` to perform the two stage least squares regression with `tsp75` as the instrumental variable. The results of this 2sls regression for the three specifications have been juxtaposed along with that when `tsp7576` was used as the instrumental variable in Table 10

```
ivregress 2sls dlhouse (dgtsp = tsp75)
outreg2 using 'imagepath'IIc.tex, ctitle(baseline) tex(pretty frag) dec
    (4) append
ivregress 2sls dlhouse (dgtsp = tsp75) 'mainctrl'
outreg2 using 'imagepath'IIc.tex, ctitle(mainctrl) keep('keepmain') tex(
    pretty frag) dec(4) append
ivregress 2sls dlhouse (dgtsp = tsp75) 'mainctrl' 'polyctrl'
outreg2 using 'imagepath'IIc.tex, title("Instrument is tsp7576 for models
     1-3 and tsp75 for models 4-6") ctitle(mainpoly) keep('keepmain' '
    keeppoly') tex(pretty frag) dec(4) append
estat endog
estat firststage
```

II c(6): Compare the findings.

From Table 10, we observe that both instruments suggest similar results though the extent of the negative effect is marginally lower when `tsp75` is used as the instrumental variable. This result is expected because `tsp7576` captures both 1975 and 1976 EPA regulated counties whereas `tsp75` only captures 1975 EPA regulated counties. Since the regulation would have naturally been enforced at a higher rate through 1976, we would expect a greater divergence between the house prices in highly polluted areas vis-a-vis less polluted areas.

Question II (d)

II d(1): Provide a concise and coherent summary of your results.

Table 10

Instrument is tsp7576 for models 1-3 and tsp75 for models 4-6

VARIABLES	(1) baseline	(2) mainctrl	(3) mainpoly	(4) baseline	(5) mainctrl	(6) mainpoly
dgtsp	-0.0037**	-0.0022**	-0.0027***	-0.0035**	-0.0020**	-0.0021**
	(0.0014)	(0.0009)	(0.0010)	(0.0014)	(0.0009)	(0.0009)
dincome		0.0001***	0.0001***		0.0001***	0.0001***
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
dunemp		-0.8244***	-1.9775***		-0.8608***	-1.8930***
		(0.2441)	(0.4617)		(0.2486)	(0.4415)
dwhite		-0.5547***	-0.3770**		-0.5545***	-0.3822***
		(0.0977)	(0.1520)		(0.0976)	(0.1478)
dpoverty		-0.6876***	-0.8305**		-0.6513***	-0.6957*
		(0.1769)	(0.3821)		(0.1776)	(0.3751)
downer		-0.0946	-0.4811**		-0.1154	-0.4700**
		(0.0975)	(0.2042)		(0.0985)	(0.1966)
owner2			-4.8432***			-4.8067***
			(1.7654)			(1.6978)
age2			52.4923***			52.1836***
			(14.7103)			(14.1338)
poverty2			-9.3690			-7.4924
			(7.0886)			(6.8915)
Constant	0.2453***	0.2295***	0.2771*	0.2474***	0.2329***	0.2891*
	(0.0129)	(0.0691)	(0.1598)	(0.0126)	(0.0678)	(0.1527)
Observations	1,000	995	995	975	970	970
R^2		0.4548	0.5153		0.4712	0.5646

Standard errors in parentheses

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

In order to understand the effect of air pollution on housing prices, we utilize an exogenous event (the 1975 federal clean air regulation) to disentangle the inherent endogeneity in modeling house prices as a function of air pollution levels and other economic indicators. We utilize the instrumental variable approach, where the instrumental variable used is the dummy indicating whether a county was regulated by the EPA in the year 1975 or in the years 1975 and 1976. Clearly this instrument has the qualities of of being correlated with the levels of air pollution but not directly correlated with housing prices. Table 10 demonstrates our findings that pollution levels are indeed negatively correlated with house prices. Our results are robust even when main economic control variables are included, and when both main and polynomials of main economic indicators are used as control variables. Finally, we find that both `tsp75` and `tsp7576` serve as good instruments and yield similar results.

II d(2): Discuss the "credibility" of the research designs underlying the results.

As discussed above, we demonstrate that the instruments chosen satisfy the required conditions. Secondly we also demonstrate that our results are consistent across the three specifications and over the two instrumental variables used. Specifically we are able to ascertain a shift in pollution levels but not a shift in house prices because of the exogenous event that was the 1975 clean air regulation. With the data, as well as with the argument, we conclude that our results are robust.