Recitation Notes 4

Konrad Menzel

October 6, 2006

1 Grouped Data and Instrumental Variables

1.1 How does OLS work?

As usual, we start with our standard regression model

$$Y_i = \alpha + X_i\beta + \varepsilon_i$$

where the X_is are scalars. Let's just recall what OLS actually does

$$\hat{\beta}_{LS} = \frac{\frac{1}{N}\sum_{i=1}^{N} Y_i(X_i - \bar{X})}{\frac{1}{N}\sum_{i=1}^{N} X_i(X_i - \bar{X})} = \frac{\mathbb{E}_N[XY] - \mathbb{E}_N[X]\mathbb{E}_N[Y]}{\mathbb{E}_N[X^2] - \mathbb{E}_N[X]^2} \stackrel{N\to\infty}{\longrightarrow} \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\mathbb{E}[X^2] - \mathbb{E}[X]^2}$$

where $\mathbb{E}_N[\cdot]$ is a shorthand for the sample mean. So what exactly in this line of reasoning tells us that we have to estimate the different moments in that formula from the same data set? - the answer to this question is that this can't be seen in the formula because it's not true. From what we know about the probability limits (i.e. we can interchange the limit operation with addition, multiplication, division - and actually any other continuous transformation), we could in principle replace each sample average in that expression with anything that plims to the right value. This opens up a wide range of possibilities, so we could e.g.

- completely ignore problems with data which are missing at random (i.e. independent of the values of X_i and Y_i), and just estimate the moments off the complete data we have
- estimate the moments from entirely different data sources as long as we are confident that they cover the same population and have the same sampling design.
- we could put together "pseudo-panels" by combining independent cross sections for different points in time.
- use moments from other sources that are aggregated at a group level, e.g. $\mathbb{E}[X_iY_i|\text{group } G_i]$, and then apply the law of iterated expectation by summing over the marginal distribution $\mathbb{P}(\text{group } G_i)$ which gives us back the unconditional expectation (this is sometimes a great option if we can't access microdata about a particular variable - mention Josh's paper on the Vietnam lottery).
- ...and much more along these lines

The whole trick is that we just have to look for some way of taking (possibly conditional) expectations that just average out the noise in the data by a law of large numbers.

1.2 Group Means as Proxy Variables

Sometimes we are completely missing any microdata on a variable we would like to include in our analysis, but instead there's some aggregate data on that variable we'd like to use. More specifically, say we have a microdata set without reliable information on wages X_i^*. Suppose we observe instead some covariates W_i (for instance dummies for type of work, industry, and specific metropolitan area, but would have to include all other regressors in the equation we want to estimate), and we are able to obtain average wages for each relevant group from a different source. Then we might suspect after the previous discussion that using these group averages as proxies should be a source of attenuation bias (but the error $(X_i^* - \mathbb{E}[X_i|W_i])$ would have conditional mean zero, and therefore continue to be "classical"). It turns out that this isn't so (at least in the limit) since the reliability ratio

$$\lambda = \frac{\text{Cov}\left(\mathbb{E}[X_i|W_i], X_i^*\right)}{\text{Var}\left(\mathbb{E}[X_i|W_i]\right)} = \frac{\text{Cov}\left(\mathbb{E}[X_i|W_i], \mathbb{E}[X_i^*|W_i]\right)}{\text{Var}\left(\mathbb{E}[X_i|W_i]\right)} = \frac{\text{Cov}\left(\mathbb{E}[X_i^*|W_i], \mathbb{E}[X_i^*|W_i]\right)}{\text{Var}\left(\mathbb{E}[X_i^*|W_i]\right)} = 1$$

so that there is no bias from measurement error in X_i. The intuitive reason for this is that using group means actually converts our estimation problem into a regression of group averages, for which the individual "measurement error" averages out after conditioning on W_i. We can also interpret this regression as an instrumental variables estimator which uses the group characteristics W_i as instruments, and therefore deals with the measurement error problem in X_i.

1.3 Main types of Bias in OLS

Let's now turn to the three main situations in which OLS is biased

1. Measurement Error
2. Omitted Variables
3. Reverse Causality/Simultaneity

Now recall that in the case of classical measurement error in the regressor, i.e. $X_i = X_i^* + \nu$, the probability limit of the OLS coefficient is

$$\text{plim}_N \hat{\beta}_{LS} = \lambda\beta = \frac{\text{Cov}(X, X^*)}{\text{Var}(X^*) + \text{Var}(\nu)}\beta$$

whereas omitted variables bias in a model

$$Y_i = \alpha + X_i^*\beta + W_i\gamma + \varepsilon_i$$

for which we don't include W in the OLS regression, the LS estimator plims to

$$\text{plim}_N \hat{\beta}_{LS} = \beta + \frac{\text{Cov}(X_i^*, W_i)}{\text{Var}(X_i^*)}\gamma$$

In both cases, there is a random component in Y or X which makes us trouble. We'd therefore like to find a way of averaging out the errors like in the best-case OLS model without smoothing out the "signal" about the regressor we are mainly interested in. It turns out that this is exactly what instrumental variables do.

1.4 Grouped-Data IV

Assume we also observe a discrete variable Z which divides the sample into r groups (i.e. one group for which $Z = 1$, another with $Z = 2$, etc.). Then if we first estimate group-wise means $\mathbb{E}[Y|Z], \mathbb{E}[X|Z]$, and then regress the conditional means on another, we get a new linear estimator

$$\tilde{\beta} = \frac{\text{Cov}(\mathbb{E}_N[Y|Z], \mathbb{E}_N[X|Z])}{\text{Var}(\mathbb{E}_N[X|Z])} = \frac{\text{Var}(\mathbb{E}_N[X^*|Z])\beta + \text{Cov}(\mathbb{E}_N[X^*|Z], \mathbb{E}_N[W|Z])\gamma}{\text{Var}(\mathbb{E}_N[X^*|Z]) + \text{Var}(\mathbb{E}_N[\nu|Z])}$$

Therefore, if the grouping implicit in Z is

- independent ν and W, i.e. $\mathbb{E}[\nu|Z] = \mathbb{E}[\nu]$ and $\mathbb{E}[W|Z] = \mathbb{E}[X^*]$ (exclusion restriction), but
- contains a signal about X^*, i.e. $\text{Var}(\mathbb{E}[X|Z]) > 0$ (first stage)

we have in the limit

$$\begin{aligned} \text{Var}(\mathbb{E}[\nu|Z]) &\overset{Indep.}{=} \text{Var}(\mathbb{E}[\nu]) = 0 \\ \text{Cov}(\mathbb{E}[X^*|Z], \mathbb{E}[W|Z]) &\overset{Indep.}{=} \text{Cov}(\mathbb{E}[X^*|Z], \mathbb{E}[W]) = 0 \end{aligned}$$

so that the bias goes away as the sample gets large. Note again that as in the previous section, our estimation problems are resolved by a law of large numbers which, smoothes out the noise which was making trouble in standard OLS once we take group-wise means.
However, this also comes along with a loss of information about the within-group correlations which are also informative about the coefficient we are looking for, since the conditional variance identity (also known as the ANOVA identity) tells us that

$$\text{Var}(\mathbb{E}[X^*|Z]) = \text{Var}(X^*) - \mathbb{E}[\text{Var}(X^*|Z)] < \text{Var}(X^*)$$

We saw that for panel data, fixed-effects estimators neutralize the bias from heterogeneity *between groups* (where "groups" usually consist of multiple observations of the individual over time) by identifying parameters off the *within variation.* Instrumental variables estimators construct groups which are defined by the instruments in a way that we can get rid of the bias from *within-group* heterogeneity (e.g. in measurement error or omitted characteristics) by estimating effects only taking into account the *between-group* variation.

1.5 Two-Stage Least Squares (2SLS)

In general, we could (and typically would) use linear projections of X on Z rather than conditional expectations, but this wouldn't change the essence of the previous argument. This leads us to Two-Stage Least Squares (2SLS) which consists of two steps:

1. First Stage: regress X on the instruments Z and recover fitted values $\hat{X}_i := P_Z X_i = Z_i \hat{\pi}$, where $\hat{\pi} = (\mathbf{Z'Z})^{-1}\mathbf{Z'X}$ are the OLS coefficients from a regression of X on Z.
2. Second Stage: regress Y on $\hat{X}$, and the coefficient $\hat{\beta}_{IV}$ on $\hat{X}$ will be the 2SLS estimate for β.

Therefore, if X_i is a scalar we have

$$\hat{\beta}_{2SLS} = \frac{\text{Cov}_N(\hat{X}_i, Y_i)}{\text{Var}(\hat{X}_i)}$$

If there is only one instrument Z_i, this can be rewritten as

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \frac{\text{Cov}_N(\hat{X}_i, Y_i)}{\text{Var}(\hat{X}_i)} = \frac{\text{Cov}_N(Z_i\hat{\pi}, Y_i)}{\text{Var}(Z_i\hat{\pi})} \\
&\overset{(*)}{=} \frac{\hat{\pi}\text{Cov}(Z_i, Y_i)}{\hat{\pi}^2\text{Var}(Z_i)} \overset{OLS}{=} \frac{\frac{\text{Cov}(Z_i,X_i)}{\text{Var}(Z_i)}\text{Cov}(Z_i, Y_i)}{\left(\frac{\text{Cov}(Z_i,X_i)}{\text{Var}(Z_i)}\right)^2 \text{Var}(Z_i)} \\
&= \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)}
\end{aligned}$$

where "(*)" uses the fact that $\hat{\pi}$ is constant across observations and can therefore be pulled out of the covariances. You can convince yourself that this implies that in the one-dimensional case, 2SLS is the same as regressing each X (first stage) and Y (reduced form) on the instrument Z, and taking the ratio of the two slope parameters.
From this, we can see that the conditions on the instrument for getting a consistent estimator are

1. $\varepsilon_i := Y_i - X_i\beta - \alpha$ is uncorrelated with Z_i, which gives us $\text{Cov}(Z_i, Y_i) = \text{Cov}(Z_i, X_i)\beta$.
2. Z is correlated with X, i.e. $\text{Cov}(Z_i, X_i) \neq 0$.

If Z only takes values 0 or 1, $\mathbb{E}[Z_i] = \mathbb{P}(Z = 1)$, and the estimator simplifies further to

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)} \\
&\overset{(**)}{=} \frac{\left(\mathbb{E}_N[Y|Z=1] - \mathbb{E}_N[Y|Z=0]\right)\text{Var}(Z)}{\left(\mathbb{E}_N[X|Z=1] - \mathbb{E}_N[X|Z=0]\right)\text{Var}(Z)} \\
&= \frac{\mathbb{E}_N[Y|Z=1] - \mathbb{E}_N[Y|Z=0]}{\mathbb{E}_N[X|Z=1] - \mathbb{E}_N[X|Z=0]}
\end{aligned}$$

since

$$\begin{aligned}
(**) \quad \text{Cov}(Z_i, Y_i) &= \mathbb{E}[Z_iY_i] - \mathbb{E}[Z_i]\mathbb{E}[Y_i] = \left(\frac{\mathbb{E}[Z_iY_i]}{\mathbb{P}(Z_i=1)} - \mathbb{E}[Y_i]\right)\mathbb{P}(Z_i=1) \\
&= \Big\{\mathbb{E}[Y_i|Z_i=1] - \Big(\mathbb{E}[Y_i|Z_i=1]\mathbb{P}(Z_i=1) + \mathbb{E}[Y_i|Z_i=0][1-\mathbb{P}(Z_i=1)]\Big)\Big\}\mathbb{P}(Z_i=1) \\
&= \left(\mathbb{E}[Y_i|Z_i=1] - \mathbb{E}[Y_i|Z_i=0]\right)\mathbb{P}(Z_i=1)[1-\mathbb{P}(Z_i=1)] \\
&= \left(\mathbb{E}[Y_i|Z_i=1] - \mathbb{E}[Y_i|Z_i=0]\right)\text{Var}(Z_i)
\end{aligned}$$

The new expression for 2SLS for a binary instrument is also known as the *Wald* instrumental variables estimator.

1.6 2SLS as a Grouped Data IV (not covered in recitation)

Now assume that we have k discrete instrumental variables Z, and we can think of them as dividing your sample into r cells. Without loss of generality, we can therefore assume that the instruments are indicator functions for the R *mutually exclusive* groups, i.e.

$$Z_{ri} := \begin{cases} 1 & \text{if observation } i \text{ falls into cell } r \\ 0 & \text{otherwise} \end{cases}$$

In vector notation, we can therefore write the matrix of all instruments as

$$\mathbf{Z} = [\mathbf{Z_1}, \mathbf{Z_2}, \ldots, \mathbf{Z_R}] = \begin{bmatrix} \iota_{N_1} & 0 & 0 & \ldots & 0 \\ 0 & \iota_{N_2} & 0 & \ldots & 0 \\ 0 & 0 & \iota_{N_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \iota_{N_R} \end{bmatrix}$$

possibly after reordering the observations, where ι_n is an n-dimensional column vector of ones, and N_r is the number of observations that falls into the rth cell.
For 2SLS, the first stage consists of fitting the endogenous right-hand side variable X onto the Zs in order to obtain (again in matrix notation)

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{P_Z X} = \mathbf{Z(Z'Z)^{-1}Z'X} \\ &= \begin{bmatrix} \iota_{N_1} & 0 & 0 & \ldots & 0 \\ 0 & \iota_{N_2} & 0 & \ldots & 0 \\ 0 & 0 & \iota_{N_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \iota_{N_R} \end{bmatrix} \begin{bmatrix} N_1 & 0 & 0 & \ldots & 0 \\ 0 & N_2 & 0 & \ldots & 0 \\ 0 & 0 & N_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & N_R \end{bmatrix}^{-1} \begin{bmatrix} \iota_{N_1} & 0 & 0 & \ldots & 0 \\ 0 & \iota_{N_2} & 0 & \ldots & 0 \\ 0 & 0 & \iota_{N_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \iota_{N_R} \end{bmatrix}' \mathbf{X} \\ &= \frac{1}{N_1}\mathbf{Z_1 Z_1' X} + \frac{1}{N_2}\mathbf{Z_2 Z_2' X} + \ldots + \frac{1}{N_R}\mathbf{Z_R Z_R' X} = \begin{bmatrix} \iota_{N_1}\bar{X}_1 \\ \iota_{N_2}\bar{X}_2 \\ \vdots \\ \iota_{N_R}\bar{X}_R \end{bmatrix} = \Big[\mathbb{E}_N[X_i|Z_i] \Big]_{i=1}^{N} \end{aligned}$$

since inside the inverse, $\iota_n' \iota_n = \sum_{j=1}^{n} 1 = n$. Notice that the whole trick is that groups are mutually exclusive, so that the inner product of the instrument matrix with itself is diagonal. The interpretation of what is going on here is also straightforward: fitting to/projecting onto the group dummies is exactly the same thing as replacing the variable with group-wise means.
In the second stage, we do plain OLS of the left-hand-side endogenous variable Y on the fitted right-hand-side endogenous variable, $\hat{X}$, which gives us

$$\hat{\beta}_{2SLS} = \frac{\text{Cov}(\hat{X}, Y)}{\text{Var}(\hat{X})} = \frac{\text{Cov}(\mathbf{P_Z} X, Y)}{\text{Var}(\mathbf{P_Z} X)} = \frac{\text{Cov}(\mathbb{E}_N[X|Z], \mathbb{E}_N[Y|Z])}{\text{Var}(\mathbb{E}_N[X|Z])}$$

For the pairwise Wald IV between groups j and k, we basically do the same thing, only that we ignore all other groups, which is equivalent to doing 2SLS using instruments $\mathbf{Z_{jk}} = [\mathbf{Z_j}, \mathbf{Z_k}]$, but since now we only have one binary instrument, the estimator simplifies significantly to

$$\hat{\beta}_{jk} = \frac{\mathbb{E}_N[Y|Z_j = 1] - \mathbb{E}_N[Y|Z_k = 1]}{\mathbb{E}_N[X|Z_j = 1] - \mathbb{E}_N[X|Z_k = 1]} = \frac{\bar{Y}_j - \bar{Y}_k}{\bar{X}_j - \bar{X}_k}$$

The asymptotic variance of this estimator under homoskedasticity of ε_i, i.e. $\text{Var}(\varepsilon_i|Z_i) = \text{Var}(\varepsilon_i) = \sigma^2$ can be computed as

$$\text{Var}(\hat{\beta}_{jk}) = \frac{\text{Var}(\bar{Y}_j - \bar{Y}_k)}{(\bar{X}_j - \bar{X}_k)^2} = \frac{\sigma^2 \left(\frac{1}{N_j} + \frac{1}{N_k} \right)}{(\bar{X}_j - \bar{X}_k)^2} = \frac{\sigma^2 (N_j + N_k)}{N_j N_k (\bar{X}_j - \bar{X}_k)^2}$$

If we write out the 2SLS estimator, we get

$$
\begin{aligned}
\hat{\beta}_{2SLS} &= \frac{\mathrm{Cov}(\mathbb{E}_N[X|Z], \mathbb{E}_N[Y|Z])}{\mathrm{Var}(\mathbb{E}_N[X|Z])} \\
&= \frac{\sum_{r=1}^R N_r \bar{X}_r (\bar{Y}_r - \bar{Y})}{\sum_{r=1}^R N_r \bar{X}_r (\bar{X}_r - \bar{X})} \\
&= \frac{\sum_{r=1}^R N_r \bar{X}_r \sum_{s=1}^R \frac{N_s}{N} (\bar{Y}_r - \bar{Y}_s)}{\sum_{r=1}^R N_r \bar{X}_r (\bar{X}_r - \bar{X})} \\
&= \frac{\sum_{r=1}^R \sum_{s>r} N_r N_s (\bar{X}_r - \bar{X}_s)(\bar{Y}_r - \bar{Y}_s)}{N \sum_{r=1}^R N_r \bar{X}_r (\bar{X}_r - \bar{X})} \\
&= \sum_{r=1}^R \sum_{s>r} \frac{N_r N_s (\bar{X}_r - \bar{X}_s)^2}{N \sum_{t=1}^R N_t \bar{X}_t (\bar{X}_t - \bar{X})} \hat{\beta}_{rs} =: \sum_{r=1}^R \sum_{s=r+1}^R w_{rs} \hat{\beta}_{rs}
\end{aligned}
$$

Notice that under homoskedasticity of ε_i, the weights on the pairwise Wald estimates $\hat{\beta}_{rs}$ in the 2SLS estimator are

$$
w_{rs} = \frac{N_r N_s (\bar{X}_r - \bar{X}_s)^2}{N \sum_{t=1}^R N_t \bar{X}_t (\bar{X}_t - \bar{X})} \propto \frac{N_r + N_s}{\mathrm{Var}(\hat{\beta}_{rs})}
$$

where "$\propto$" stands for "proportional to". This is exactly what the statement in the Angrist (1991) paper means that 2SLS is an "efficient GLS-combination" of the pairwise Wald-IV estimators.

1.7 Potential Problems: Weak Instruments, Many Instruments

As we can see from the derivation, even with "clean" instruments (i.e. IVs for which the exclusion restrictions hold) there are other potential problems in finite samples. As we saw above, IV works well under two conditions:

1. Taking expectations conditional on Z averages out the measurement error in X or omitted variables in Y - "spunk" in Jerry's terminology - by the law of large numbers
2. Smoothing conditional on Z doesn't average out the "signal" in the explanatory variable

However, if the sample is small, the IV estimator may fail to do this properly for several reasons:

- If the cells defined by Z in the grouped-data IV remain relatively small even in a large sample, taking group-wise averages doesn't get rid of much of the "spunk". As an extreme example, one could argue that a dummy variable for each observation are valid instruments (the (ex ante) expectation of measurement error is certainly zero for each individual, and X will probably differ from one observation to another). But the first stage would in this case give us a perfect fit, and the fitted values are going to be identical to the actual variable, so that 2SLS would do exactly the same thing as OLS. Having many instruments may biase the 2SLS estimator towards OLS due to "overfitting" of the first stage.
- If $\mathrm{Var}(\mathbb{E}[X|Z]) \approx 0$, we may with some probability end up with a sample for which $\mathbb{E}_N[X_i|Z_i = 1] - \mathbb{E}_N[X_i|Z_i = 0]$ is arbitrarily close to zero. Since this is the denominator of the corresponding Wald IV, this will blow up the estimator, and we are likely to end up with very extreme estimates of either sign. You can also see from the 2SLS formula that the same thing will happen there if the first stage is very weak. The finite-sample expectation and variance of a 2SLS with one

instrument and one endogenous regressor are actually not defined (even though Stata will still give you the asymptotic standard errors). However the median of 2SLS will still get it right, so that if we have one instrument and one endogenous variable, this is only an issue about the precision of the estimator.

- If both problems come together - i.e. if we have a very small first stage and more instrumental variables than we actually need to identify the model - the weak instruments exacerbate any bias from correlations between the endogenous variables which are still picked up by the instrument. An approximation to the bias can be given as[1]

$$Bias \approx \frac{(k-2)\sigma_{\varepsilon,\nu}}{NR^2\sigma_X^2}$$

where $\nu_i := X_i - Z_i\pi$ is first stage residual, k is the number of instrumental variables, and R^2 is the unadjusted first-stage R-squared. Therefore, if instruments are sufficiently weak (i.e. the first-stage R-squared is very low), the bias may be very severe even if the number of instruments is small relative to the sample size so that overfitting is not really an issue. This is often referred to as the *weak instruments problem.* Intuitively, IV fails to perform well because the instruments pick up "more noise than signal" from the endogenous variable.

So it's always good to look at the first stage of your 2SLS regression (to check whether it's sufficiently large and has a plausible sign), and be suspicious if your estimate with many IVs is much closer to OLS than if you just use as many instruments as necessary to identify the parameters. However, this section was only supposed to point out some potential problems, and you'll see this in greater depth in 14.382.

2 Discussion of the Replication of Angrist (1991)

While all of you got most of the replication exercise right, I'd just like to go over Problem 3(b) on the first problem set again, because it's a nice illustration of some of the main issues in the empirics of life-cycle labor supply.

OLS Regressions

The OLS coefficients are pretty close to Angrist (1991), and significantly negative. As we said before, this doesn't make sense. We'd suspect that there are two sources of bias: marginal utility of lifetime wealth, λ, is likely going to be lower for individuals with generally high wages, which would cause a negative omitted variables bias. In addition, and more importantly, we generated wages by dividing earnings by hours, which is likely to lead to division bias, which is also negative.

```
. reg lnh lnw, robust

Regression with robust standard errors                  Number of obs
=   15829
                                                        F(  1, 15827) =  142.64
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.0172
                                                        Root MSE      =  .30717
```

[1] Bound Jaeger and Baker (1995): Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak, JASA *90*

```
------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |   -.074059    .006201   -11.94   0.000    -.0862136   -.0619043
       _cons |   7.763484    .008386   925.77   0.000     7.747047    7.779922
------------------------------------------------------------------------------

. reg lnh lnw year, robust

Regression with robust standard errors                Number of obs
=   15829
                                                      F(  2, 15826) =   77.29
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0183
                                                      Root MSE      =  .30701

------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.0723591   .0061878   -11.69   0.000    -.0844879   -.0602303
        year |  -.0032761   .0008112    -4.04   0.000     -.004866   -.0016861
       _cons |   8.003794   .0602887   132.76   0.000     7.885622    8.121967
------------------------------------------------------------------------------

. reg lnh lnw year year2, robust

Regression with robust standard errors                Number of obs
=   15829
                                                      F(  3, 15825) =   53.41
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0190
                                                      Root MSE      =  .30691

------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.0728822   .0061873   -11.78   0.000    -.0850099   -.0607545
        year |   .1322667    .043383     3.05   0.002     .0472311    .2173022
       year2 |  -.0009158   .0002935    -3.12   0.002     -.001491   -.0003406
       _cons |    2.99822   1.600729     1.87   0.061    -.1393911    6.135832
------------------------------------------------------------------------------
```

Fixed-Effects Regressions

You should remember at this point that the main motivation for the fixed-effects approach at this point is that in the dynamic programming problem under perfect foresight, the (individual-specific) Lagrange multiplier does not only capture all information about wages and prices in other periods, but is also constant over time (by definition). In the MaCurdy specification, lambda-constant labor supply is multiplicatively separable in the current wage and the Lagrange multiplier, so that after taking logs, the Lagrange multiplier is absorbed by an additive individual fixed effect.
My ANACOVA/Fixed-Effects estimates are again similar to those in Angrist (1991) and even more negative than OLS. From our discussion of panel data, you should remember that the fixed-effects transformation exacerbates the problems with measurement error. I.e. FE might be even more negative than OLS because the division bias is under sensible assumptions about serial correlation even more severe in the within-regression. Also if there's no longer perfect foresight about wages, the Lagrange multipliers might not only change over time, but they'd be negatively correlated with wages. This would invalidate the fixed-effects strategy, and we'd get negative omitted variables bias if the true intertemporal elasticity of substitution is positive (which it should be).

```
. xtreg lnh lnw, fe i(id)

Fixed-effects (within) regression               Number of obs      =
15829 Group variable (i): id                          Number of
groups   =      1439

R-sq:  within  = 0.0939                         Obs per group: min =
11
       between = 0.0002                                        avg =      11.0
       overall = 0.0172                                        max =        11

                                                F(1,14389)         =   1491.79
corr(u_i, Xb)  = -0.4558                        Prob > F           =
0.0000

------------------------------------------------------------------------------
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.2627622   .0068031   -38.62   0.000    -.2760972   -.2494272
       _cons |   7.998834   .0087009   919.31   0.000     7.981779    8.015889
-------------+----------------------------------------------------------------
     sigma_u |  .22734189
     sigma_e |  .24242703
         rho |  .46792124   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1438, 14389) =     7.66         Prob >
F = 0.0000

. xtreg lnh lnw year, fe i(id)

Fixed-effects (within) regression               Number of obs      =
15829 Group variable (i): id                          Number of
```

```
groups   =      1439

R-sq:  within  = 0.0940                         Obs per group: min =
11
       between = 0.0002                                        avg =      11.0
       overall = 0.0173                                        max =        11

                                                F(2,14388)         =    745.98
corr(u_i, Xb)  = -0.4547                        Prob > F           =
0.0000

------------------------------------------------------------------------------
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.2621607   .0069094   -37.94   0.000     -.275704   -.2486174
        year |  -.0003086   .0006188    -0.50   0.618    -.0015217    .0009044
       _cons |   8.020923   .0451372   177.70   0.000     7.932448    8.109398
-------------+----------------------------------------------------------------
     sigma_u |  .22719147
     sigma_e |  .24243336
         rho |  .46757869   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1438, 14388) =     7.64         Prob >
F = 0.0000

. xtreg lnh lnw year year2, fe i(id)

Fixed-effects (within) regression               Number of obs      =
15829 Group variable (i): id                          Number of
groups   =      1439

R-sq:  within  = 0.0963                         Obs per group: min =
11
       between = 0.0002                                        avg =      11.0
       overall = 0.0177                                        max =        11

                                                F(3,14387)         =    511.06
corr(u_i, Xb)  = -0.4583                        Prob > F           =
0.0000

------------------------------------------------------------------------------
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.2650943   .0069173   -38.32   0.000    -.2786531   -.2515355
        year |   .1975371   .0323412     6.11   0.000     .1341441    .2609301
       year2 |  -.0013365   .0002184    -6.12   0.000    -.0017646   -.0009083
       _cons |   .7159575   1.194753     0.60   0.549    -1.625913    3.057828
-------------+----------------------------------------------------------------
```

```
     sigma_u |  .22792748
     sigma_e |  .24212696
         rho |  .46981937   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1438, 14387) =     7.68         Prob >
F = 0.0000
```

2SLS Estimates

2SLS should definitely kill the measurement error, and also the omitted variables bias if the instruments are actually exogenous. One thing we can see right away is that it is much more difficult to replicate the IV estimates since standard errors are high despite the relatively large sample size.
The exclusion restriction on the instruments is that, after controlling for a linear or quadratic time trend, the lifetime income effects that enter labor supply decisions through the marginal utility average out in the aggregate for each period - in other words, conditional on the parametric trend, all remaining variation in the marginal utility of wealth is cross-sectional. As you can see from the first stages of the IV regressions, the linear time trend is collinear with the full set of year dummies, so that one IV has to be dropped (similarly, for the quadratic trend, we have to drop 2 exogenous variables from the first stage, and so on). So in total, for a k^{th} order approximation to the time trend, we have $11 - k$ instruments to identify the ISE, so that the equation is overidentified by a degree of $11 - k - 1$.
The results support the measurement error hypothesis because for all three specifications of the time trend, IV estimates are significantly higher than FE coefficients. This constitutes the basis of the Angrist (1991) test for measurement error.

```
. ivreg lnh (lnw=yr*), robust first note: yr11 dropped due to
collinearity

First-stage regressions
-----------------------

      Source |       SS       df       MS              Number of obs =   15829
-------------+------------------------------           F( 10, 15818) =   15.96
       Model |  47.6370661    10  4.76370661           Prob > F      =  0.0000
    Residual |  4721.96357 15818  .298518369           R-squared     =  0.0100
-------------+------------------------------           Adj R-squared =  0.0094
       Total |  4769.60063 15828  .301339438           Root MSE      =  .54637

------------------------------------------------------------------------------
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         yr1 |  -.1653508    .020369    -8.12   0.000    -.2052764   -.1254252
         yr2 |  -.1236961    .020369    -6.07   0.000    -.1636217   -.0837705
         yr3 |  -.1010749    .020369    -4.96   0.000    -.1410005   -.0611493
         yr4 |  -.0447062    .020369    -2.19   0.028    -.0846318   -.0047806
         yr5 |  -.0230886    .020369    -1.13   0.257    -.0630142     .016837
         yr6 |   -.028766    .020369    -1.41   0.158    -.0686916    .0111596
         yr7 |  -.0435892    .020369    -2.14   0.032    -.0835148   -.0036636
         yr8 |  -.0109082    .020369    -0.54   0.592    -.0508338    .0290174
         yr9 |   .0090843    .020369     0.45   0.656    -.0308413    .0490099
```

```
        yr10 |   .0051704    .020369     0.25   0.800    -.0347552     .045096
       _cons |   1.295098   .0144031    89.92   0.000     1.266867     1.32333
------------------------------------------------------------------------------


IV (2SLS) regression with robust standard errors       Number of obs
=   15829
                                                       F(  1, 15827) =    8.45
                                                       Prob > F      =  0.0036
                                                       R-squared     =  0.0060
                                                       Root MSE      =  .30893

------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |  -.1339499     .04607    -2.91   0.004    -.2242524   -.0436474
       _cons |    7.83818   .0575247   136.26   0.000     7.725425    7.950935
------------------------------------------------------------------------------
Instrumented:  lnw Instruments:   yr1 yr2 yr3 yr4 yr5 yr6 yr7 yr8
yr9 yr10
------------------------------------------------------------------------------

. ivreg lnh year (lnw=yr*), robust first note: yr11 dropped due to
collinearity

First-stage regressions
-----------------------

      Source |       SS       df       MS              Number of obs =   15829
-------------+------------------------------           F( 10, 15818) =   15.96
       Model |  47.6370661    10  4.76370661           Prob > F      =  0.0000
    Residual |  4721.96357 15818  .298518369           R-squared     =  0.0100
-------------+------------------------------           Adj R-squared =  0.0094
       Total |  4769.60063 15828  .301339438           Root MSE      =  .54637

------------------------------------------------------------------------------
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .0165351   .0020369     8.12   0.000     .0125425    .0205276
         yr1 |  (dropped)
         yr2 |   .0251196   .0194308     1.29   0.196     -.012967    .0632062
         yr3 |   .0312058   .0186685     1.67   0.095    -.0053867    .0677982
         yr4 |   .0710394   .0181044     3.92   0.000     .0355527     .106526
         yr5 |   .0761219   .0177573     4.29   0.000     .0413156    .1109282
         yr6 |   .0539094   .0176401     3.06   0.002     .0193328    .0884859
         yr7 |   .0225511   .0177573     1.27   0.204    -.0122552    .0573575
         yr8 |   .0386971   .0181044     2.14   0.033     .0032104    .0741837
```

```
         yr9 |   .0421544   .0186685     2.26   0.024      .005562    .0787469
        yr10 |   .0217055   .0194308     1.12   0.264    -.0163811    .0597921
       _cons |  -.0111727   .1510744    -0.07   0.941    -.3072958    .2849504
------------------------------------------------------------------------------

IV (2SLS) regression with robust standard errors       Number of obs
=   15829
                                                       F(  2, 15826) =   12.67
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .44074

------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |   .5060088    .146741     3.45   0.001     .2183798    .7936379
        year |  -.0123186   .0026191    -4.70   0.000    -.0174522   -.0071849
       _cons |     7.9516   .0849728    93.58   0.000     7.785043    8.118156
------------------------------------------------------------------------------
Instrumented:  lnw Instruments:   year yr1 yr2 yr3 yr4 yr5 yr6 yr7
yr8 yr9 yr10
------------------------------------------------------------------------------

. ivreg lnh year year2 (lnw=yr*), robust first note: yr11 dropped
due to collinearity

First-stage regressions
-----------------------

      Source |       SS       df       MS              Number of obs =   15829
-------------+------------------------------           F( 10, 15818) =   15.96
       Model |  47.6370661    10  4.76370661           Prob > F      =  0.0000
    Residual |  4721.96357 15818  .298518369           R-squared     =  0.0100
-------------+------------------------------           Adj R-squared =  0.0094
       Total |  4769.60063 15828  .301339438           Root MSE      =  .54637

------------------------------------------------------------------------------
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |  (dropped)
       year2 |   .0001117   .0000138     8.12   0.000     .0000847    .0001387
         yr1 |  (dropped)
         yr2 |   .0261251   .0194831     1.34   0.180    -.0120639    .0643141
         yr3 |   .0329933   .0187417     1.76   0.078    -.0037426    .0697293
         yr4 |   .0733856   .0181716     4.04   0.000     .0377672     .109004
         yr5 |   .0788033   .0177982     4.43   0.000     .0439168    .1136898
```

yr6	.0567024	.0176434	3.21	0.001	.0221193	.0912856
yr7	.0252325	.0177224	1.42	0.155	-.0095055	.0599705
yr8	.0410433	.0180415	2.27	0.023	.0056798	.0764068
yr9	.043942	.0185976	2.36	0.018	.0074886	.0803954
yr10	.022711	.0193792	1.17	0.241	-.0152744	.0606964
_cons	.5978321	.0763914	7.83	0.000	.4480963	.7475679

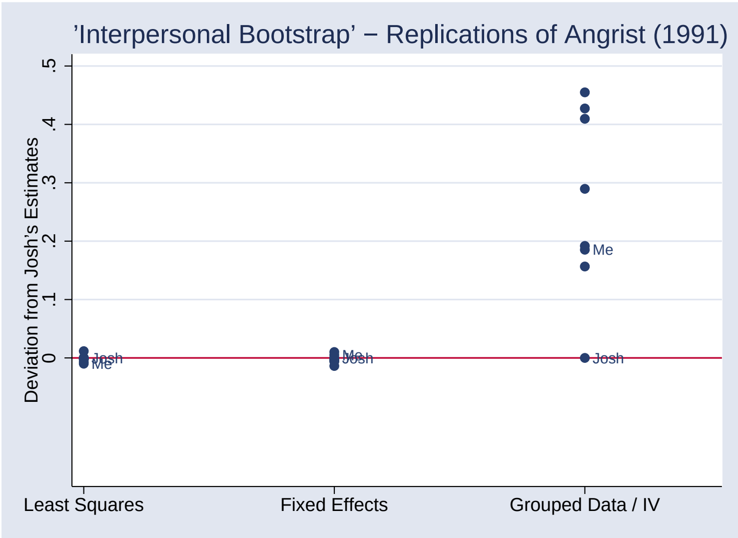
IV (2SLS) regression with robust standard errors Number of obs = 15829

F(3, 15825) = 5.69
Prob > F = 0.0007
R-squared = .
Root MSE = .57608

lnh	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnw	.8193102	.3251283	2.52	0.012	.1820218	1.456599
year	-.1706996	.137047	-1.25	0.213	-.4393272	.097928
year2	.001037	.0008984	1.15	0.248	-.000724	.0027981
_cons	13.59182	4.883424	2.78	0.005	4.019749	23.16388

Instrumented: lnw Instruments: year year2 yr1 yr2 yr3 yr4 yr5 yr6 yr7 yr8 yr9 yr10

If I plot all replication results from the problem set answers you handed in (specification 2 with a quadratic trend) - including my own - you can see that due to sample size, OLS and Fixed-Effects estimates were relatively stable with regard to how you put together the data set applied the sample selection criteria. However, 2SLS isn't just much less precisely estimated according to the asymptotic standard errors in the Stata output, but it's also very sensitive to the definition of the sample, as you can see from the figure. I would also like to point out that no two people of us all got exactly the same sample or identical results for any of the specifications.



The Card Critique

One more thing to notice about the 2SLS estimates is that the better we control for time trends, the higher our estimate of the intertemporal elasticity of substitution becomes. This fits very well with the omitted variables bias story for the FE regression from not taking into account changes in λ_i that we could - under reasonable assumptions - expect to be *negatively* correlated with present wages (Why?). So let's see what happens if we allow for a cubic time trend:

```
IV (2SLS) regression with robust standard errors        Number of obs
=   15829
                                                        F(  4, 15824) =    9.36
                                                        Prob > F      =  0.0000
                                                        R-squared     =       .
                                                        Root MSE      =  .64194

------------------------------------------------------------------------------
             |               Robust
         lnh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnw |   .9589432   .4019281     2.39   0.017     .1711184    1.746768
        year |  -3.375522   4.043445    -0.83   0.404    -11.30113     4.55009
       year2 |   .0440566   .0542894     0.81   0.417    -.0623567    .1504699
```

```
       year3 |  -.0001924    .000243    -0.79   0.428    -.0006686    .0002838
       _cons |   92.96352   100.1449     0.93   0.353    -103.3319    289.2589
------------------------------------------------------------------------------
Instrumented:  lnw Instruments:   year year2 year3 yr1 yr2 yr3 yr4
yr5 yr6 yr7 yr8 yr9 yr10
------------------------------------------------------------------------------
```

...and again, we managed to increase our estimate of the ISE by another 14 percentage points. This is basically the main critique of this approach in the Card (1994) paper: if the true problem is omitted variables bias from *aggregate* trends in shocks to lifetime wealth (more precisely: shocks to *expected* lifetime wealth), time as an instrument doesn't satisfy an exclusion restriction. Now, the data we are using for this exercise is from the 1970s which is the textbook example for unexpected macro shocks (in a few buzz words: oil crisis, stagflation, disappearance of the Phillips curve etc.). From our theory that the income effect of an increase in lifetime wealth on labor supply should be negative, we'd therefore interpret these IV coefficients which do not control sufficiently for aggregate time trends as underestimating the true ISE.