

# 14.661: Recitation 3

Chris Walters

October 3, 2010

## 1 OLS Problems

Recall that if we estimate the bivariate regression

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

we get

$$plim \hat{\beta}_{OLS} = \frac{Cov(x_i, y_i)}{Var(x_i)}$$

There are some important cases where OLS fails to estimate a parameter of interest.

### 1.1 Omitted Variable Bias

Suppose the regression we want to run is

$$y_i = \alpha + \beta x_i + \gamma z_i + \eta_i$$

with  $Cov(x_i, \eta_i) = Cov(z_i, \eta_i) = 0$ . However, we can't observe  $z_i$ , so we omit it from our regression and instead run:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

In this case, the probability limit of our OLS estimator is

$$\begin{aligned} plim \hat{\beta}_{OLS} &= \frac{Cov(x_i, y_i)}{Var(x_i)} \\ &= \frac{Cov(x_i, \alpha + \beta x_i + \gamma z_i + \eta_i)}{Var(x_i)} \end{aligned}$$

$$\Rightarrow plim \hat{\beta}_{OLS} = \beta + \gamma \cdot \frac{Cov(x_i, z_i)}{Var(x_i)}$$

You should remember this formula. Note that the bivariate OLS estimator is consistent if  $z_i$  has a coefficient of zero in the full regression, or if  $x_i$  and  $z_i$  aren't correlated.

## 1.2 Measurement Error

Suppose we want to run the regression

$$y_i^* = \alpha + \beta x_i^* + \epsilon_i$$

with  $Cov(x_i^*, \epsilon_i) = 0$ . We observe  $y_i^*$ , but instead of observing  $x_i^*$  we instead observe

$$x_i = x_i^* + v_i$$

with  $Cov(v_i, x_i^*) = Cov(v_i, \epsilon_i) = 0$ . What happens if we regress  $y_i^*$  on  $x_i$ ? Note that we can write

$$y_i^* = \alpha + \beta x_i + (\epsilon_i - \beta v_i)$$

Since  $v_i$  shows up in both the error term and in our regressor, we are in trouble. The result of running this regression is

$$\begin{aligned} plim \hat{\beta}_{OLS} &= \frac{Cov(x_i, y_i^*)}{Var(x_i)} \\ &= \frac{Cov(x_i^* + v_i, \alpha + \beta x_i^* + \epsilon_i)}{Var(x_i)} \\ &= \beta \cdot \frac{Var(x_i^*)}{Var(x_i)} \end{aligned}$$

Under the conditions assumed above this is

$$\boxed{\implies plim \hat{\beta}_{OLS} = \beta \cdot \frac{Var(x_i^*)}{Var(x_i^*) + Var(v_i)}}$$

The quantity

$$\lambda = \frac{Var(x_i^*)}{Var(x_i^*) + Var(v_i)}$$

is called the “reliability ratio;”  $\lambda$  is the proportion of the variance in the observed  $x_i$  due to the variance of the true variable of interest (“signal” rather than “noise”). Since  $\lambda \in (0, 1)$ , we have

$$|plim \hat{\beta}_{OLS}| < |\beta|$$

That is, random measurement error causes our parameter estimate to be too close to zero. This is called “attenuation bias.” Note that if you are only interested in testing the null hypothesis that  $\beta = 0$ , then if you can reject this hypothesis the possibility of measurement error strengthens your conclusion.

## 2 Instrumental Variables (IV)

### 2.1 Basic IV

Let's go back to the bivariate equation

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Suppose we are worried that  $x_i$  is correlated with  $\epsilon_i$ , either because of OVB or measurement error. The general solution is Instrumental Variables (IV). To do IV, we need to find a variable  $z_i$ , called an instrument, with the following properties:

1.  $Cov(z_i, \epsilon_i) = 0$ ; this is “exogeneity” or the “exclusion restriction”. This restriction is inherently untestable.
2.  $Cov(z_i, x_i) \neq 0$ ; this is called “first stage” or “relevance”. Note that this is testable.

Suppose we have a  $z_i$  that we would like to use. The IV estimator is

$$\hat{\beta}_{IV} = \frac{\widehat{Cov}(y_i, z_i)}{\widehat{Cov}(x_i, z_i)}$$

so

$$plim \hat{\beta}_{IV} = \frac{Cov(y_i, z_i)}{Cov(x_i, z_i)}$$

If  $z_i$  satisfies assumptions (1) and (2), then

$$\begin{aligned} plim \hat{\beta}_{IV} &= \frac{Cov(\alpha + \beta x_i + \epsilon_i, z_i)}{Cov(x_i, z_i)} \\ &= \beta \cdot \frac{Cov(x_i, z_i)}{Cov(x_i, z_i)} + \frac{Cov(\epsilon_i, z_i)}{Cov(x_i, z_i)} \\ &= \beta \end{aligned}$$

Note that

$$\frac{Cov(y_i, z_i)}{Cov(x_i, z_i)} = \frac{Cov(y_i, z_i)/Var(z_i)}{Cov(x_i, z_i)/Var(z_i)}$$

From this equation we can see that the IV estimator is really the ratio of two regression coefficients. Consider the two equations

$$y_i = \gamma_0 + \gamma_1 z_i + u_i$$

$$x_i = \pi_0 + \pi_1 z_i + w_i$$

Then

$$\hat{\beta}_{IV} = \frac{\hat{\gamma}_1}{\hat{\pi}_1}$$

The first equation is called the “reduced form;” this is the regression of  $y_i$  on the instrument  $z_i$ . The second equation is called the “first stage;” it is the regression of  $x_i$  on  $z_i$ . The instrumental variables estimator is the ratio of the reduced form and first stage coefficients. This is intuitive. If we find that  $\gamma_1 \neq 0$ , under the assumptions above this can only be because of  $x$ ; since  $z_i$  is correlated with  $x$  but uncorrelated with anything else that determines  $y$ , the correlation between  $y$  and  $z$  must be due to a relationship between  $x$  and  $z$ . IV attributes the relationship between  $y$  and  $z$  to  $x$ , adjusting the relationship between  $x$  and  $z$  to determine the effect of a 1-unit change in  $x$ .

## 2.2 The Wald Estimator

The Wald estimator is a famous special case of the single-instrument setup above were  $z_i$  is a dummy variable:  $z_i \in \{0, 1\}$ . In this case, it is straightforward to show that the IV estimand becomes

$$\frac{Cov(y_i, z_i)}{Cov(x_i, z_i)} = \frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{E[x_i|z_i = 1] - E[x_i|z_i = 0]}$$

The Wald estimator is the sample analogue of this quantity. Note that if  $x_i$  is also a dummy variable, the denominator simply becomes the increase in the probability of “treatment” caused by  $z_i$  switching from 0 to 1.

## 2.3 Multiple Instruments: 2SLS

Suppose we have many variables  $z_1, z_2, \dots, z_K$  that we think satisfy conditions (1) and (2) above. How do we choose which one to use? We could compute  $\hat{\beta}_{IV}^k$  separately for each instrument. But we would like to somehow combine the instruments to produce our “best” estimate of  $\beta$ . Two-stage least squares (2SLS) is a way of doing this. To do 2SLS, we perform the following procedure:

1. Run  $x_i = \pi_0 + \sum_k \pi_k z_{ik} + u_i$ . As before, this regression of  $x_i$  on the instruments is called the “first stage.” From the first stage estimates, produce the fitted values  $\hat{x}_i$ .
2. Run the “second stage” regression  $y_i = \alpha + \beta \hat{x}_i + \epsilon_i$ ;  $\hat{\beta}_{2SLS}$  is the estimate of  $\beta$  from this equation.

The second stage regression is just bivariate OLS, so we know that

$$plim \hat{\beta}_{2SLS} = \frac{Cov(y_i, \hat{x}_i)}{Var(\hat{x}_i)}$$

For intuition, note that the second stage is a regression of  $y_i$  on the part of  $x_i$  that is explained by the instruments. We can think of this as the “good” part of  $x_i$ ; it is the part that is uncontaminated by correlation with  $\epsilon_i$ .

Furthermore, note that since  $Cov(x_i, \hat{x}_i) = Cov(\tilde{x}_i + \hat{x}_i, \hat{x}_i) = Var(\hat{x}_i)$ , we can write this as

$$plim \hat{\beta}_{2SLS} = \frac{Cov(y_i, \hat{x}_i)}{Cov(x_i, \hat{x}_i)}$$

2SLS is therefore univariate IV, where the instrument is the fitted values from the first stage. We would like to do IV using some function of our many instruments; 2SLS selects the linear projection of  $x$  onto the  $z$ ’s. In 382, you will see that this is the efficient linear combination of instruments under homoskedasticity.

One last note on 2SLS: Don’t do it manually! You will get the wrong standard errors. Instead, you should just let Stata do both stages together.

## 2.4 Multiple Endogenous Variables

Now suppose our equation of interest is

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

and we are worried that both  $x_1$  and  $x_2$  are endogenous (correlated with  $\epsilon_i$ ). As long as we have at least two instruments, this is no problem; we can do 2SLS just like before. We run a first stage for each endogenous variable:

$$x_{1i} = \pi_{10} + \sum_k \pi_{1k} z_{ik} + u_{1i}$$

$$x_{2i} = \pi_{20} + \sum_k \pi_{2k} z_{ik} + u_{2i},$$

get fitted values  $\hat{x}_1$  and  $\hat{x}_2$ , and run the second-stage equation

$$y_i = \alpha + \beta_1 \hat{x}_{1i} + \beta_2 \hat{x}_{2i} + \epsilon_i$$

## 2.5 IV and Grouped Regression

Finally, an interesting special case of IV. Suppose that we want to estimate

$$y_i = \alpha + \beta x_i + \epsilon_i$$

and we have instruments  $z_1 \dots z_K$ , which are mutually exclusive and exhaustive dummy variables. This is the situation in Angrist (1991), where the  $z$ 's are year dummies. Suppose that  $n_k$  observations have  $z_{ik} = 1$ .

We want to use these instruments to run 2SLS. The first stage is

$$x_i = \sum_k \pi_k z_{ik} + u_i$$

What are the fitted values from this regression? As we saw in the case of Fixed Effects, the fitted values from a regression on an exclusive and exhaustive set of group dummies are just the group means. That is,

$$\hat{x}_i = \bar{x}_k,$$

where  $k$  is the group that  $i$  is in. So in our data, the fitted values look like this:

$$\hat{x} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_1 \\ \vdots \\ \bar{x}_K \\ \vdots \\ \bar{x}_K \end{bmatrix}$$

We have  $n_1$  copies of  $\bar{x}_1$ , followed by  $n_2$  copies of  $\bar{x}_2$ , etc.

Next, we want to run the second stage. Before doing this, we should note the following useful property of fitted values:

$$Cov(y_i, \hat{x}_i) = Cov(\hat{y}_i + \tilde{y}_i, \hat{x}_i) = Cov(\hat{y}_i, \hat{x}_i)$$

where  $\hat{y}$  and  $\tilde{y}$  are fitted values and residuals from a regression of  $y$  on the  $z$ 's. Therefore, in the second stage, we can put the  $\hat{y}$ 's on the left-hand side instead of the  $y$ 's and it doesn't make a difference. The second stage OLS regression is therefore

$$\bar{y}_k = \alpha + \beta \bar{x}_k + \epsilon_i,$$

where there are  $n_k$  copies of each  $(\bar{y}_k, \bar{x}_k)$  pair in the data. This is a regression of group means on group means, with  $n_k$  copies of each pair. Therefore,

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}_{2SLS} \\ \hat{\beta}_{2SLS} \end{pmatrix} &= \arg \min_{\alpha, \beta} \sum_i (\bar{y}_k - \alpha - \beta \bar{x}_k)^2 \\ \implies \begin{pmatrix} \hat{\alpha}_{2SLS} \\ \hat{\beta}_{2SLS} \end{pmatrix} &= \arg \min_{\alpha, \beta} \sum_k n_k (\bar{y}_k - \alpha - \beta \bar{x}_k)^2 \end{aligned}$$

This is a weighted least squares (WLS) regression using the group-means, with weights equal to group sizes. This is GLS on the equation

$$\bar{y}_k = \alpha + \beta \bar{x}_k + \epsilon_k$$

So, to interpret: IV with group dummies as instruments is equivalent to GLS estimation of the group-mean equation. Note the contrast with fixed effects. When we do FE, we assume that permanent variation across groups is "contaminated" and therefore use only the within-group variation. When we do IV, we assume that the within-group variation is contaminated and use only the variation in means across groups.

Finally, one more fact: The minimized GLS minimand is

$$\hat{T} = \sum_k \frac{n_k}{\hat{\sigma}^2} \cdot (\bar{y}_k - \hat{\alpha} - \hat{\beta} \bar{x}_k)^2$$

This statistic doubles as an overidentification test of the validity of the instruments. Under the null hypothesis that all of the instruments are valid, it will have a  $\chi^2_{K-1}$  distribution. (We can do an overid test whenever we have more instruments than endogenous variables). If it is larger than the relevant critical value from the  $\chi^2$  distribution, we will reject the hypothesis that all of our instruments are valid, though we will not know which ones are invalid.

## 3 Life Cycle Model

### 3.1 Model recap

In class we saw the LC model

$$V(p, w, A_0) = \max_{c_1 \dots c_T, h_1 \dots h_T} \sum_t \frac{1}{(1 + \rho)^t} \cdot [\gamma_1 c_t^{\delta_1} - \gamma_2 h_t^{\delta_2}]$$

s.t.

$$\sum_t \frac{p_t c_t}{(1 + r)^t} \leq A_0 + \sum_t \frac{w_t h_t}{(1 + r)^t}$$

Just looking at the problem setup, we can already make some important observations.

1. In this problem there are no constraints on savings; in a given period,  $s_t = w_t h_t - p_t c_t$  can be anything, positive or negative. In other words, the agent can borrow and lend freely as much as she wants at the interest rate  $r$ . You can imagine other versions of this problem with constraints on savings (for example  $s_t \geq 0$ , no borrowing), or with different interest rates for borrowing and for lending, etc. Such extensions will often importantly change the conclusions we get from the model.
2. Concavity of the within period utility function (here we will get this by assuming  $\delta_2 > 1$ ,  $\delta_1 < 1$ ) gives an incentive to smooth consumption and work effort over time. It doesn't make sense for the agent to consume a lot in some periods and a little in others – since she can freely move resources through time, she will want to equalize her marginal utilities from consumption in different time periods. The same incentive exists for hours, though in analyzing this model we usually focus on fluctuations in the wage rate over time and assume goods prices are roughly fixed. This will generate a countervailing incentive to work more in some periods and less in others, as we will see.
3. Using the envelope theorem and assigning  $\lambda$  as the Lagrange multiplier on the constraint, we can see that

$$\frac{\partial V}{\partial A_0} = \lambda$$

$\lambda$  can therefore be interpreted as the marginal utility the agent gets from another dollar in present value; it is her “lifetime” marginal utility of income.  $V$  will be concave in  $A$ , so as the agent gets richer  $\lambda$  goes down. When we solve the maximization problem,  $\lambda$  will end up being a function of all wages, prices, and the non-labor income level.

For simplicity set  $p_t \equiv 1 \forall t$ . If we set up a Lagrangian, take FOCs, and then take logs, we will obtain

$$\log c_t = -\frac{\log \lambda}{(1 - \delta_1)} - \frac{(\rho - r)}{(1 - \delta_1)} t + \frac{\log \gamma_1 + \log \delta_1}{(1 - \delta_1)}$$

$$\log h_t = \frac{\log \lambda}{(\delta_2 - 1)} + \frac{\log w_t}{(\delta_2 - 1)} + \frac{(\rho - r)}{(\delta_2 - 1)} t - \frac{(\log \gamma_2 - \log \delta_2)}{(\delta_2 - 1)}$$

The main insights of the lifecycle model are visible from these two equations.

1.  $c_t$  only depends on  $w_t$  through  $\lambda$ .  $\lambda$  is a sufficient statistic for wages in all periods when it comes to consumption. To put it another way, current consumption is completely divorced from current period wages; all that matters is the agent's lifetime marginal utility of income.
2. The effect of a change in  $w_t$  on  $h_t$  operates through two channels.

$$\frac{\partial \log h_t}{\partial \log w_t} = \delta + \delta \cdot \frac{\partial \log \lambda}{\partial \log w_t}$$

where  $\delta \equiv \frac{1}{\delta_2 - 1} > 0$ . We call  $\delta$  the Intertemporal Substitution Elasticity (ISE). This equation tells us that the effect of a wage change has a substitution effect (the ISE), and then an effect that operates through the change in  $\lambda$ ; this is a lifetime wealth effect. Note that these two effects act in opposite directions. A higher wage means that the agent is better off, so her marginal utility will go down as the wage in a given period increases.

This model therefore offers a simple framework for thinking about how changes in wages should affect labor supply. The ISE is the relevant parameter for evaluating the effects of wage changes that do not affect the marginal utility of income. Such changes are either small and temporary (so that to a first approximation they do not affect  $\lambda$ ) or anticipated (so that they are already built into  $\lambda$ ). As we've written the problem, it doesn't really make sense to talk about anticipated vs. unanticipated changes since we've assumed perfect certainty, but the basic insights go through if we add uncertainty and rational expectations. In contrast, any change that is large, permanent, and unanticipated (like getting fired from a job for which you've accumulated a lot of firm-specific human capital) should have an effect on  $\lambda$ . The ISE will therefore be the wrong parameter for evaluating the effects of changes like this.

### 3.2 Lifecycle metrics

We've spent some time in lecture talking about how to estimate the parameters of the lifecycle model. In particular, if we have data on individuals  $i$ , the labor supply equation from section 3.1 become

$$\log h_{it} = \delta \log w_{it} + \delta(\rho - r)t + F_i + \eta_{it}$$

How do we estimate this equation?

#### 1. Approach 1: Ordinary Least Squares

- Problem 1:  $F_i$  is unobservable and includes  $\lambda_i$ , which will be negatively correlated with  $\log w_{it}$ . From the omitted variables bias formula, we know that this will tend to make our estimates of  $\delta$  too small.

#### 2. Approach 2: Fixed Effects

- Fixed effects will solve the problem with omitting  $F_i$  since it is constant over time. This is the approach taken by a lot of labor supply papers from the 1980s.
- Problem 2: But wages tend to be measured with error, and doing fixed effects exacerbates measurement error problems. In fact, both OLS and FE on this equation are vulnerable to a special type of measurement error known as "division bias," which can even reverse the sign of the coefficient of interest. Indeed, many old labor supply papers find very small and even negative estimates of  $\delta$ . I'll discuss this more in section 3.3.

#### 3. Approach 3: Instrumental Variables

- The general solution to omitted variable bias and measurement error is IV. Angrist (1991) estimates the panel labor supply equation using time dummies as instruments. The rationale for this is that cyclical shocks associated with the business cycle should be interpreted as shifts in labor demand; these shocks are not large enough to alter  $\lambda$ , so there should be no year effects on labor supply independent from changes in the wage. As we saw in section 2.5, IV using time dummies as instruments is algebraically equivalent to estimating the grouped equation

$$\log \bar{h}_t = \delta \log \bar{w}_t + \delta(\rho - r)\bar{t} + \bar{F} + \bar{\eta}_t$$



- Angrist (1991) estimates this grouped equation and gets an estimate of  $\delta$  around 0.6, which is much larger than the usual FE estimates and suggests that the lifecycle model may be useful for explaining variation in labor supply.
- Problem 3: The exclusion restriction required for the Angrist (1991) approach to work may not hold. In particular, Card (1994) argues that shocks associated with the business cycle may be large enough to substantially alter lifetime wealth levels, in which case there should be time effects in the above equation ( $\lambda$  should change with the business cycle) and the grouped regression will not identify the ISE.

#### 4. Approach 4: Short-term experiments

- Problems 1-3 have left us without any airtight estimates of  $\delta$ . The modern labor supply literature uses credibly exogenous short-term variation in wages coming from randomized experiments (e.g. the Fehr and Goette paper on bike messengers) or specific industries with special characteristics (e.g. the Oettinger paper on stadium vendors or the Camerer et al. paper on cab drivers) to achieve identification of the ISE.
- These papers tend to find fairly large estimates of the ISE (though this is not always true in the cab drivers literature)
- Problem 4: These short-term experiments may have better identification than traditional labor supply studies, but they look only at very small, idiosyncratic professions. The workers and jobs in these professions may have characteristics that differ importantly from those of the general population, and these may affect their labor supply behavior. For example, bike messengers are generally very young. Furthermore, they have the opportunity to choose when and how long they work. While this may help with identification of the ISE for this group, most jobs do not have this feature. To put this critique another way, these experiments may have good internal validity, but their external validity is not clear.

### 3.3 Division Bias

It is worth going through the math of division bias a little more carefully. Suppose we are not worried about OVB for the moment, and we want to estimate the equation

$$\log h_i^* = \alpha + \delta \log w_i^* + \eta_i$$

If we had data on true hours and wages  $h_i^*$  and  $w_i^*$  we would be able to consistently estimate  $\delta$ ; in other words,  $Cov(\log w_i^*, \eta_i) = 0$ .

However, many people do not report an explicit hourly wage, so we have to compute  $w$ . If we had perfect measures of hours and earnings, we could successfully compute the true  $w_i^*$ :

$$w_i^* = \frac{y_i^*}{h_i^*}$$

Furthermore, hours are measured with multiplicative error  $v_i$ , which we can assume is independent of everything else in the model. We observe:

$$h_i = h_i^* \cdot v_i, y_i^*$$

So observed wages are

$$w_i = \frac{y_i^*}{h_i} = \frac{y_i^*}{v_i h_i^*}$$

$$\implies \log w_i = \log w_i^* - \log v_i$$

Suppose we regress observed hours on observed wages using OLS. Then we obtain

$$\begin{aligned} plim \hat{\delta} &= \frac{Cov(\log h_i, \log w_i)}{Var(\log w_i)} \\ &= \frac{Cov(\log h_i^* + v_i, \log w_i^* - \log v_i)}{Var(\log w_i^* - \log v_i)} \\ &= \frac{Cov(\alpha + \delta \log w_i^* + \eta_i + \log v_i, \log w_i^* - \log v_i)}{Var(\log w_i^* - \log v_i)} \\ &= \delta \frac{Var(\log w_i^*)}{Var(\log w_i^*) + Var(\log v_i)} - \frac{Var(\log v_i)}{Var(\log w_i^*) + Var(\log v_i)} \end{aligned}$$

The first term shows the usual attenuation bias result  $-\delta$  is multiplied by a positive number less than one, so the measurement error in  $w$  will pull our estimate towards zero. However, we also have a second term that is unambiguously negative; this term comes from the correlation between the measurement error on the left hand side and the measurement error in the denominator of our right hand side variable of interest. With a positive  $\delta$ , this makes attenuation bias worse and can even reverse the sign of the coefficient.

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.661 Labor Economics I  
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.