

## Econ 311: Ability Bias

Consider the model

$$\log y_{it} = \alpha_0 + \alpha_1 S_i + U_{it},$$

where  $y_{it}$  = income,  $S_i$ , = schooling, and  $\alpha_0$  and  $\alpha_1$  are parameters of interest. What we have omitted from the above specification is unobserved ability, which is captured in the residual term  $U_{it}$ . We thus re-write the above as

$$\log y_{it} = \alpha_0 + \alpha_1 S_i + a_i + \varepsilon_{it}$$

where  $a_i$  is ability,  $(\varepsilon_{it}, \varepsilon_{i't'}) \perp\!\!\!\perp (S_i, S_{i'})$ , and

we believe that  $\text{Cov}(a_i, S_i) \neq 0$ . Thus,

$E(U_{it} | S_i) \neq 0$ , so that OLS on our original specification gives biased and inconsistent estimates.

## **1.1 Strategies for Estimation**

**IV.** Find proxies for ability and include them as regressors. Examples may include: height, weight, etc. The problem with this approach is that proxies may measure ability with error and thus introduce additional bias (see 1.3).

**Fixed Effect Method.** Find a paired comparison. Examples may include a genetic twin or sibling with similar or identical ability.

Consider two individuals  $i$  and  $i'$ :

$$\begin{aligned}\log y_{it} - \log y_{i't} &= (\alpha_0 + \alpha_1 S_i + U_{it}) \\ &\quad - (\alpha_0 + \alpha_1 S_{i'} + U_{i't}) \\ &= \alpha_1 (S_i - S_{i'}) + (a_i - a_{i'}) + (\varepsilon_{it} - \varepsilon_{i't}).\end{aligned}$$

Note: if  $a_i = a_{i'}$ , then OLS performed on our fixed effect estimator is unbiased and consistent. If  $a_i \neq a_{i'}$ , then we just get a different bias (see 1.2). Further, if  $S_i$  is

measured with error, we may exacerbate the bias in our fixed effect estimator (see 1.3).

## 1.2. OLS vs. Fixed Effect (FE)

$$\text{plim } \alpha_1^{OLS} = \alpha_1 = \alpha_1 + \frac{\text{Cov}(a, S)}{\text{Var}(S)}$$

(See derivation of equation 2.2 for more

background on the above derivation). We also

impose

$$\text{Var}(S) = \text{Var}(S')$$

$$\text{Cov}(a, S) = \text{Cov}(a', S')$$

$$\text{Cov}(a', S) = \text{Cov}(a, S')$$

With these assumptions, our fixed effect

estimator is given by:

$$\text{plim } \alpha_1^{FE} = \alpha_1 + \frac{\text{Cov}(S-S', (a-a') + (\varepsilon-\varepsilon'))}{\text{Var}(S-S')}$$

$$= \alpha_1 + \frac{Cov(a, S) - Cov(a', S)}{Var(S) - Cov(S, S')}$$

Note that if  $Cov(a', S) = 0$ , and ability is positively correlated with schooling, then the fixed effect estimator is upward biased. From the preceding, we see that the fixed effect estimator has more asymptotic bias if:

$$\begin{aligned} \frac{Cov(a, S) - Cov(a', S)}{Var(S) - Cov(S, S')} &> \frac{Cov(a, S)}{Var(S)} \\ \Rightarrow \frac{Cov(a, S)}{Var(S)} &> \frac{Cov(a', S)}{Cov(S, S')} \end{aligned}$$

### 1.3. Measurement Error

Say  $S^* = S + \nu$ , where  $S^*$  is observed schooling. Our model now becomes

$$\log y = \alpha_0 + \alpha_1 S + U = \alpha_0 + \alpha_1 S^* + (a + \varepsilon - \alpha_1 \nu)$$

and the fixed effect estimator gives

$$\begin{aligned}
\log y - \log y' &= (\alpha_0 + \alpha_1 S + U) \\
&\quad - (\alpha_0 + \alpha_1 S' + U') \\
&= \alpha_1 (S^* - S^{*'}) + (U - U') \\
&\quad + \alpha_1 (\nu' - \nu).
\end{aligned}$$

Now we wish to examine which estimator, (*OLS* or fixed effect), has more asymptotic bias given our measurement error problem.

For the remaining arguments of this section,

we assume

$$E(\nu | S) = E(\nu' | S) = E(\nu | \nu') = 0$$

so that the *OLS* estimator gives

$$\begin{aligned}\text{plim } \alpha_1^{OLS} &= \alpha_1 + \frac{\text{Cov}(S^*, a + \varepsilon - \alpha_1 \nu)}{\text{Var}(S^*)} \\ &= \alpha_1 + \frac{\text{Cov}(a, S) - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu)}.\end{aligned}$$

The fixed effect estimator gives

$$\begin{aligned}\text{plim } \alpha_1^{FE} &= \\ &\alpha_1 + \frac{\text{Cov}(S^* - S^{*'}, (U - U') + \alpha_1(\nu' - \nu))}{\text{Var}(S^* - S^{*'})} \\ &= \alpha_1 + \frac{\text{Cov}((S - S'), (a - a')) - \alpha_1 \text{Var}(\nu' - \nu)}{\text{Var}(S - S') + \text{Var}(\nu' - \nu)} \\ &= \alpha_1 + \frac{\text{Cov}(a, S) - \text{Cov}(a, S') - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu) - \text{Cov}(S', S)}\end{aligned}$$

.

Under what conditions will the fixed effect bias be greater? From the above, we

know that this will be true if and only if

$$\begin{aligned}
& \frac{Cov(a, S) - Cov(a, S') - \alpha_1 Var(\nu)}{Var(S) + Var(\nu) - Cov(S', S)} \\
& > \frac{Cov(a, S) - \alpha_1 Var(\nu)}{Var(S) + Var(\nu)} \\
& \Rightarrow Cov(a, S') (Var(S) + Var(\nu)) \\
& > (\alpha_1 Var(\nu) - Cov(a, S)) Cov(S', S) \\
& \Rightarrow \frac{Cov(a, S) - \alpha_1 Var(\nu)}{Var(S) + Var(\nu)} > \frac{Cov(a, S')}{Cov(S', S)}.
\end{aligned}$$

If this inequality holds, taking differences can actually worsen the fit over *OLS* alone.

Intuitively, we see that we have differenced out the true component,  $S$ , and compounded our measurement error problem with the fixed

effect estimator.

## 2. Errors in Variables

### 2.1. The Model

Suppose that the equation for earnings is

given by

$$Y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + U_t,$$

where  $E(U_t | X_{1t}, X_{2t}) = 0 \forall t, t'$ . Also

define

$$X_{1t}^* = X_{1t} + \varepsilon_{1t} \text{ and } X_{2t}^* = X_{2t} + \varepsilon_{2t}.$$

Here,  $X_{1t}^*$  and  $X_{2t}^*$  are observed and measure  $X_{1t}$  and  $X_{2t}$  with error. We also impose that  $X_i \perp\!\!\!\perp \varepsilon_j \forall i, j$ . So, our initial model can be equivalently re-written as

$$Y_t = X_{1t}^* \beta_1 + X_{2t}^* \beta_2 + (U_t - \varepsilon_{1t} \beta_1 - \varepsilon_{2t} \beta_2).$$

Finally, by assumed independence of  $X$  and  $\varepsilon$ , we write

$$\sum_{x^*} = \sum_x + \sum_\varepsilon.$$

## 2.2 McCallum's Problem

Question: Is it better for estimation of  $\beta_1$  to include other variables measured with error? Suppose that  $X_{1t}$  is not measured with error, in the sense that  $\varepsilon_{1t} = 0$ , while  $X_{2t}$  is measured with error. In 2.2.1 and 2.2.2 below, we consider both excluding and including

$X_{2t}$ , and investigate the asymptotic properties of both cases.

**Excluded**  $X_{2t}$ . The equation for earnings with omitted  $X_2$  is

$$y = X_1\beta_1 + (U + X_2\beta_2).$$

Therefore, by arguments similar to those in the appendix, we know

$$(2.1) \text{plim } \tilde{\beta}_1 = \beta_1 + \frac{\sigma_{12}}{\sigma_{11}}\beta_2.$$

Here,  $\sigma_{12}$  is the covariance between the regressors, and  $\sigma_{11}$  is the variance of  $X_1$ .

Before moving on to a more general model for the inclusion of  $X_{2t}$ , let us first consider

the classical case for including both variables.

Suppose

$$\Sigma_{\epsilon} = \begin{bmatrix} \sigma_{11}^* & 0 \\ 0 & \sigma_{22}^* \end{bmatrix},$$

$$\Sigma_x = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}.$$

We know that

$$(2.2) \text{ plim } \hat{\beta} = \left[ I - \left( \Sigma_{x^*} \right)^{-1} \left( \Sigma_{\epsilon} \right) \right] \beta,$$

where the coefficient and regressor vec-

tors have been stacked appropriately (see

Appendix for derivation). Note that  $\Sigma_{\epsilon}$

represents the variance-covariance matrix

of the measurement errors, and  $\Sigma_x$  is the

variance-covariance matrix of the regressors.

Straightforward computations thus give

$$\begin{aligned}
\text{plim} \hat{\beta} &= \\
& \left[ I - \begin{bmatrix} \sigma_{11} + \sigma_{11}^* & 0 \\ 0 & \sigma_{22} + \sigma_{22}^* \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11}^* & 0 \\ 0 & \sigma_{22}^* \end{bmatrix} \right] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma_{11}}{\sigma_{11} + \sigma_{11}^*} & 0 \\ 0 & \frac{\sigma_{22}}{\sigma_{22} + \sigma_{22}^*} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.
\end{aligned}$$

Included  $X_{2t}$ . In McCallum's problem

we suppose that  $\sigma_{12}^* = 0$ . Further, as  $X_{1t}$  is

not measured with error,  $\sigma_{11}^* = 0$ . Substituting

this into equation 2.2 yields

$$\text{plim} \hat{\beta} = \beta - \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} + \sigma_{22}^* \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{22}^* \end{bmatrix} \beta$$

With a little algebra, the above gives

$$\text{plim} \hat{\beta}_1 = \beta_1 + \beta_2 \left( \frac{\sigma_{12}}{\sigma_{11}} \right) \left( \frac{\sigma_{22}^*}{\sigma_{22} + \sigma_{22}^* - \frac{\sigma_{12}^2}{\sigma_{11}}} \right)$$

$$= \beta_1 + \beta_2 \left( \frac{\sigma_{12}}{\sigma_{11}} \right) \left( \frac{\sigma_{22}^*}{\sigma_{22}(1-\rho_{12}^2) + \sigma_{22}^*} \right)$$

where  $\rho_{12}^2$  is simply the correlation coefficient,

$\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}$ . Further, we know that

$$0 < \rho_{12}^2 < 1$$

so including  $X_{2t}$  results in less asymptotic bias (inconsistency). (We get this result by comparing the above with the bias from excluding  $X_{2t}$  in section 2.2.1). So, we have justified the kitchen sink approach. This result generalizes to the multiple regressor case - 1 badly measured variable with  $k$  good ones

*(Econometrica, 1972).*

## 2.3 General Case

$$\text{plim } \hat{\beta} = \beta - (\Sigma_{x^*})^{-1} \Sigma_{\varepsilon} \beta = \beta -$$

$$\begin{bmatrix} \sigma_{11} + \sigma_{11}^* & \sigma_{12} + \sigma_{12}^* \\ \sigma_{12} + \sigma_{12}^* & \sigma_{22} + \sigma_{22}^* \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

With a little algebra we find

$$\begin{aligned} \det(\Sigma_{x^*}) &= \sigma_{11}\sigma_{22} - \sigma_{12}^2 + \sigma_{11}^*\sigma_{22}^* \\ &\quad - \sigma_{12}^{*2} + \sigma_{11}^*\sigma_{22} + \sigma_{11}\sigma_{22}^* - 2\sigma_{12}\sigma_{12}^* \end{aligned}$$

Therefore,

$$\text{plim } \hat{\beta} = \beta - \frac{1}{\det(\Sigma_{x^*})} \begin{bmatrix} \sigma_{22} + \sigma_{22}^* & -(\sigma_{12} + \sigma_{12}^*) \\ -(\sigma_{12} + \sigma_{12}^*) & \sigma_{11} + \sigma_{11}^* \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Supposing  $\sigma_{12}^* = 0$ ,

$$\det(\tilde{\Sigma}_{x^*}) = \det(\Sigma_{x^*}) \Big|_{\sigma_{12}^*=0}$$

$$= \sigma_{11}\sigma_{22} - \sigma_{12}^2 + \sigma_{11}^*\sigma_{22}^* + \sigma_{11}^*\sigma_{22} + \sigma_{11}\sigma_{22}^*$$

and thus,

$$\text{plim } \hat{\beta} = \beta - \begin{bmatrix} \frac{(\sigma_{22} + \sigma_{22}^*)\sigma_{11} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} & \frac{\sigma_{12}\sigma_{22}^*}{\det(\tilde{\Sigma}_{x^*})} \\ \frac{\sigma_{11}^*\sigma_{12}}{\det(\tilde{\Sigma}_{x^*})} & \frac{(\sigma_{11} + \sigma_{11}^*)\sigma_{22} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Note that if  $\beta_2\sigma_{12} < 0$ , OLS may not be

downward biased for  $\beta_1$ . If  $\beta_2 = 0$ ,

$$\text{plim } \hat{\beta}_2 = \frac{\beta_1\sigma_{12}\sigma_{11}^*}{\det(\tilde{\Sigma}_{x^*})},$$

so, if  $X_2$  were a race variable and blacks get

lower quality schooling, (where schooling is measured by  $X_{1t}$ ,) then  $\sigma_{12} < 0$ , and hence  $\hat{\beta}_2 < 0$ . This would be a finding in support of labor market discrimination.

## 2.4 The Kitchen Sink Revisited

McCallum's analysis suggests that one should toss in a variable measured with error if there is no measurement error in  $X_{1t}$ . But suppose that there is measurement error in  $X_{1t}$ . Is it still better to include the additional variable measured with error as a regressor? We proceed by imposing  $\beta_2 = 0$ .

Excluded  $X_{2t}$ . The equation for earnings with measurement error in  $X_1$  and excluded  $X_2$  is

$$\begin{aligned}
y &= (X_1^* + \varepsilon_1) \beta_1 + (U + X_2 \beta_2) \\
&= X_1^* \beta_1 + (U + X_2 \beta_2 + \beta_1 \varepsilon_1).
\end{aligned}$$

Therefore

$$\begin{aligned}
(2.3) \text{ plim } \tilde{\beta}_1 &= \beta_1 - \beta_1 \left( \frac{\sigma_{11}^*}{\sigma_{11} + \sigma_{11}^*} \right) \\
&= \beta_1 \left( \frac{\sigma_{11}}{\sigma_{11} + \sigma_{11}^*} \right) = \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right).
\end{aligned}$$

Included  $X_{2t}$ . From our analysis in the

General Case (Section 2.3), we know that

$$\text{plim } \hat{\beta}_1 = \beta_1 \left( \frac{(\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} \right).$$

If  $\sigma_{22}^* = 0$ , so that  $X_{2t}$  is not measured

with error,

$$\begin{aligned} \text{plim } \hat{\beta}_1 &= \beta_1 \left( \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2 + \sigma_{11}^*\sigma_{22}} \right) \\ &= \beta_1 \left( \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right). \end{aligned}$$

Comparing the two, we see that adding the variable measured with error always exacerbates the bias. For, the excluded bias will be smaller if

$$\begin{aligned} \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right) &> \beta_1 \left( \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right) \\ \text{iff } \left( 1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) &> \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) (1 - \rho_{12}^2) \\ \text{iff } 0 &> -\rho_{12}^2 \frac{\sigma_{11}^*}{\sigma_{11}} \end{aligned}$$

which is always the case, provided  $\rho_{12}^2 > 0$ .

(Note that the coefficients on  $\beta_1$  for both the excluded and included case are less than one. So, the larger coefficient is the one with less bias, as stated above.) Now suppose that  $\sigma_{22}^* > 0$ , so that both variables are measured with error. Then

$$\begin{aligned} \text{plim } \hat{\beta}_1 &= \beta_1 \left( \frac{(\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} \right) \\ &= \beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2}{1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2} \right). \end{aligned}$$

Intuitively, adding measurement error in

$X_{2t}$  can only worsen the bias, and thus exclusion should again be preferred to inclusion. Formally, including  $X_{2t}$  gives more bias if and only if

$$\beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2}{1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2} \right) < \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right)$$

$$\text{iff } \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) \left( 1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2 \right)$$

$$< \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2 \right)$$

$$\text{iff } -\rho_{12}^2 \frac{\sigma_{11}^*}{\sigma_{11}} < 0.$$

Provided  $\rho_{12}^2 > 0$ , including  $X_{2t}$  results in less bias than excluding it. If  $\rho_{12}^2 = 0$ , the bias from including  $X_{2t}$  is obviously seen to

be

$$\begin{aligned} \beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}}}{1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}}} \right) &= \beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}}}{\left(1 + \frac{\sigma_{22}^*}{\sigma_{22}}\right) \left(1 + \frac{\sigma_{11}^*}{\sigma_{11}}\right)} \right) \\ &= \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right) \end{aligned}$$

so that including and excluding  $X_{2t}$  yields the

same result. Finally, from the General Case

section, we have:

$$\begin{aligned} \text{plim } \hat{\beta}_1 \\ &= \frac{\beta_1 (\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2 + \beta_2 (\sigma_{12} \sigma_{22}^*)}{\sigma_{11} \sigma_{22} - \sigma_{12}^2 + \sigma_{11}^* \sigma_{22}^* + \sigma_{11}^* \sigma_{22} + \sigma_{11} \sigma_{22}^*}. \end{aligned}$$

L'Hopital's rule on the above shows that

$$\lim_{\sigma_{11}^* \rightarrow \infty} \left( \text{plim } \hat{\beta}_1 \right) = 0, \text{ and}$$

$$\lim_{\sigma_{22}^* \rightarrow \infty} \left( \text{plim } \hat{\beta}_1 \right) \rightarrow \frac{\beta_1 \sigma_{11} + \beta_2 \sigma_{12}}{\sigma_{11} + \sigma_{11}^*}$$

$$= \frac{\beta_1 \sigma_{11}}{\sigma_{11} + \sigma_{11}^*} + \frac{\beta_2 \sigma_{12}}{\sigma_{11} + \sigma_{11}^*}.$$

### 3. Sibling Models: Components of Variance Scheme

Suppose that data on two brothers, say  $\alpha$  and  $\beta$ , is at our disposal. Without loss of generality, we will consider how to estimate parameters of interest for person  $\alpha$  in what follows. We will begin by introducing a general model and then focus on the two-person case mentioned above. Consider the following triangular system:

$$y_{1ij} = \varepsilon_{1ij}$$

$$y_{2ij} = \nu_{12}y_{1ij} + \varepsilon_{2ij}$$

$$y_{3ij} = \nu_{13}y_{1ij} + \nu_{23}y_{2ij} + \varepsilon_{3ij}$$

Here,  $j$  indexes the  $j^{\text{th}}$  person in the  $i^{\text{th}}$  group.

We assume that  $\varepsilon_{lij}$  and  $\varepsilon_{li'j}$  are uncorrelated (i.e., uncorrelated across groups). Further, we suppose

$$\varepsilon_{kij} = \lambda_k h_{ij} + \mu_{kij}$$

$$h_{ij} = F_i + g_{ij},$$

for  $k = 1, 2, 3$ . We assume  $\mu_{kij}$  is uncorrelated across equations and across  $j$  within the group,  $F_i$  is iid across groups, and  $g_{ij}$  is iid within groups and uncorrelated with  $F_i$ .

### 3.1 Estimation

We specialize the above model into a two person framework and propose a similar three equation system. Let  $y_1 =$  early (pre-school) test score,  $y_2 =$  schooling (years), and  $y_3 =$  earnings. It seems plausible to write the equation system

$$y_1 = h + U_1$$

$$y_2 = \lambda_2 h + U_2.$$

$$y_3 = \nu_{23} y_2 + \lambda_3 h + U_3,$$

where  $h$  = ability. Regressing  $y_3$  on  $y_2$

clearly gives biased estimates of  $\nu_{23}$  as

$E(h | y_2) \neq 0$ . If  $\lambda_3 > 0$ , then OLS estimates

of  $\nu_{23}$  are upward biased. One estimation

approach is to use  $y_1$  as a proxy for ability

$$y_3 = \nu_{23} y_2 + \lambda_3 (y_1 - U_1) + U_3.$$

However, this results in a similar problem

— regressing  $y_3$  on  $y_1$  and  $y_2$  will give biased estimates as  $y_1$  is correlated with our residual. (i.e.,  $y_1$  is an imperfect proxy).

### **Solutions:**

One solution is to use  $y_{1\beta}$  as an instrument for  $y_{1\alpha}$ . Why is this a valid IV? From our construction of the model in 1.3, we know that the  $U_i$  are uncorrelated across equations and groups. Further, test scores are correlated across siblings. That is,  $Cov(y_{1\alpha}, y_{1\beta}) \neq 0$  by our group structure.

Another solution is possible if there exists an additional early reading on the same person

$$y_0 = \lambda_0 h + U_0.$$

Then if  $\lambda_0 \neq 0$ ,  $y_0$  is a valid proxy for  $y_1$ , and we can perform 2SLS.

### 3.2. Griliches and Chamberlain model

$$y_1 = \lambda_1 h + U_1$$

$$y_2 = \nu_{12} y_1 + \lambda_2 h + U_2$$

$$y_3 = \nu_{13} y_1 + \nu_{23} y_2 + \lambda_3 h + U_3$$

where  $y_1$  = years schooling,  $y_2$  = late test score (SAT), and  $y_3$  = earnings. Note that there are alternative models with other dependent variables. For example,  $\{y_1 =$  schooling,  $y_2 =$  early earnings, and  $y_3 =$  late earnings $\}$ , and  $\{y_1 =$  schooling,  $y_2 =$

consumption, and  $y_3 = \text{earnings}$ ). Getting the equation system into reduced form and expressing as matrix notation, we write

$$y_k = d_k h + \rho_k,$$

where

$$d_k = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 + \nu_{12}\lambda_1 \\ \lambda_3 + \nu_{13}\lambda_1 + \nu_{23}(\lambda_2 + \nu_{12}\lambda_1) \end{bmatrix}$$

and

$$\rho_k = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 + \nu_{12}\mu_1 \\ \mu_3 + \nu_{13}\mu_1 + \nu_{23}(\mu_2 + \nu_{12}\mu_1) \end{bmatrix}.$$

**Estimation.** For estimation, we impose

that  $\nu_{23} = 0$ . In our second example of section 3.2, this would be equivalent to stating

that there is no correlation between transient income and consumption (permanent income hypothesis). In general, with one factor, we need one more exclusion than that implied by triangularity.

i.  $y_1$  proxies  $h$ .

$$h = \frac{y_1 - \rho_1}{d_1}$$

so that

$$y_2 = \frac{d_2}{d_1}y_1 - \frac{d_2}{d_1}\rho_1 + \rho_2.$$

We can then estimate  $\frac{d_2}{d_1}$  consistently by using  $y_{1\beta}$  as an instrument for  $y_{1\alpha}$  in the equation above.

ii. Get residuals from  $I$ :  $z = \rho_2 - \frac{d_2}{d_1}\rho_1$ .

iii. Use the residuals as an instrument for  $y_1$  in the  $y_3$  equation.  $Z$  is valid since it is

both uncorrelated with  $h$  and  $U_3$ , and it is correlated with  $y_1$ :

$$\begin{aligned} Cov(y_1, z) &= Cov\left(y_1, \rho_2 - \frac{d_2}{d_1}\rho_1\right) \\ &= Cov\left(\lambda_1 h + U_1, U_2 + \nu_{12}U_1 - \frac{\lambda_2 + \nu_{12}\lambda_1}{\lambda_1}U_1\right) \\ &= Cov\left(\lambda_1 h + U_1, U_2 - \frac{\lambda_2}{\lambda_1}U_1\right) \neq 0 \end{aligned}$$

if  $U_1 \neq 0$ , and,  $\lambda_2 \neq 0$ . Thus we can estimate  $\nu_{13}$ .

iv. Interchange the role of  $y_2$  and  $y_3$  to estimate  $\nu_{12}$ .

v. Form the residual (and recall that  $\nu_{13}$  is known and  $\nu_{23} = 0$ )

$$w = y_3 - \nu_{13}y_1 = \lambda_3 h + U_3.$$

vi. Use  $y_1$  as a proxy for ability. Subbing this into  $V$  gives

$$w = \frac{\lambda_3}{\lambda_1}y_1 + \frac{\lambda_3}{\lambda_1}U_3 - \frac{\lambda_3}{\lambda_1}U_1.$$

vii. Now use  $y_{1\beta}$  as an instrument for  $y_{1\alpha}$

in the above to get an estimate of  $\frac{\lambda_3}{\lambda_1}$ .  
viii. Interchange the role of  $y_2$  and  $y_3$  to estimate  $\frac{\lambda_2}{\lambda_1}$ .

### **3.3 Triangular systems more generally**

Without loss of generality, suppose that  $y_2$  is excluded from the  $t^{th}$  equation of our system. (We are supposing the existence of an extra exclusion than that implied by triangularity). We seek to estimate the parameters of the system in equation  $t$  as well as equations before and after  $t$ .

## Equation t.

i. Use  $y_1$  as a proxy for ability. Solving for  $h$  and subbing into the equation

$$y_k = d_k h + \rho_k,$$

we get

$$y_k = \frac{d_k}{d_1} y_1 - \frac{d_k}{d_1} \rho_1 + \rho_k,$$

and we are considering  $k = 2, \dots, t - 1$ . The

ratio  $\frac{d_k}{d_1}$  can then be identified using  $y_{1\beta}$  as an

instrument for  $y_{1\alpha}$ .

ii. Form the residuals

$$z_k = \rho_k - \frac{d_k}{d_1} \rho_1 \quad k = 2, \dots, t - 1.$$

Now we have  $t - 2$  IV's ( $z_2, z_3, \dots, z_{t-1}$ )

for the  $t - 2$  independent variables in the

$t^{th}$  equation (  $y_1, y_3, \dots, y_{t-1}$ ), so we can consistently estimate the coefficients in the  $t^{th}$  equation.

Equations before t.

iii. Form

$$y_t^* = y_t - \nu_{1t}y_1 - \dots - \nu_{t-1t}y_{t-1}.$$

We can use  $y_1, \dots, y_{k-1}, y_t^*$  to form  $k - 1$  purged IV's and  $y_{t\beta}^*$  is used as a proxy for unobserved ability,  $h$ . In this way, we can estimate all of the parameters in equations  $k < t$ . (Note the sequential order implicit in this triangular system. We must first estimate

$t$  before this step can be made.)

**Example.** Suppose  $t > 3$  and

$$y_3 = \nu_{13}y_1 + \nu_{23}y_2 + \lambda_3h + U_3.$$

Use  $y_t^* = \lambda_t h + U_t$  as a proxy for  $h$ .

Substituting this into our  $y_3$  equation yields

$$y_3 = \nu_{13}y_1 + \nu_{23}y_2 + \frac{\lambda_3}{\lambda_t}y_t^* + \left( U_3 - \frac{\lambda_3}{\lambda_t}U_t \right).$$

Observe that  $y_1$ , and  $y_2$ , are independent of our residual, but  $y_t^*$  is not. We can use  $y_{t\beta}^*$  as an instrument for  $y_{t\alpha}^*$  to estimate the parameters above. This obviously generalizes for all equations less than  $t$ .

**Equations after  $t$ .**

iv. Assume identification for all equations through  $t$  via an exclusion restriction in equation  $t$ .

Example. As an example, consider the following:

$$y_4 = \nu_{14}y_1 + \nu_{24}y_2 + \nu_{34}y_3 + \lambda_4h + U_4$$

Define

$$y_2^* = y_2 - \nu_{12}y_1, y_3^* = y_3 - \nu_{13}y_1 - \nu_{23}y_2$$

Solving for  $y_1$  and  $y_2$  and subbing into the equation for  $y_4$ , we find

$$\begin{aligned} y_4 &= \nu_{14}y_1 + \nu_{24}y_2 \\ &+ \nu_{34}(y_3^* + \nu_{13}y_1 + \nu_{23}y_2) + \lambda_4h + U_4 \\ &= (\nu_{14} + \nu_{34}\nu_{13})y_1 + (\nu_{24} + \nu_{34}\nu_{23})y_2 \end{aligned}$$

$$\begin{aligned}
& + \nu_{34}y_3^* + \lambda_4h + U_4 \\
& = (\nu_{14} + \nu_{34}\nu_{13})y_1 + (\nu_{24} + \nu_{34}\nu_{23})(y_2^* + \nu_{12}y_1) \\
& + \nu_{34}y_3^* + \lambda_4h + U_4 \\
& = \nu_{14}^*y_1 + \nu_{24}^*y_2^* + \nu_{34}y_3^* + \lambda_4h + U_4
\end{aligned}$$

where

$$\begin{aligned}
\nu_{14}^* & = \nu_{14} + \nu_{24}^*\nu_{12} + \nu_{34}\nu_{13}, \nu_{24}^* \\
& = \nu_{24} + \nu_{23}\nu_{34}
\end{aligned}$$

Using  $y_1$  as a proxy for  $h$  and substituting

we get,

$$y_4 = \pi_1 y_1 + \nu_{24}^* y_2^* + \nu_{34} y_3^* + \left( U_4 - \frac{\lambda_4}{\lambda_1} U_1 \right)$$

where  $\pi_1 = \nu_{14}^* + \frac{\lambda_4}{\lambda_1}$ . We can then use

$y_{1\beta}$ ,  $y_{2\alpha}^*$ , and  $y_{3\alpha}^*$  as instruments to get an

estimate of  $\nu_{34}$ . Define

$$\tilde{y}_4 = y_4 - \nu_{34}y_3 = \nu_{14}y_1 + \nu_{24}y_2 + \lambda_4h + U_4$$

(Excluding  $y_3$  allows us to estimate the remaining parameters). Using  $y_3^*$  as a proxy for  $h$  yields

$$y_4 = \nu_{14}y_1 + \nu_{24}y_2 + \frac{\lambda_4}{\lambda_3}y_3^* + \left( U_4 - \frac{\lambda_4}{\lambda_3}U_3 \right).$$

We can then estimate  $\nu_{14}$ , and  $\nu_{24}$  by using  $y_{1\alpha}$ ,  $y_{2\alpha}$ , and  $y_{3\beta}^*$  as an IV. We can continue estimating. For example, consider the 5<sup>th</sup> equation:

- i. Rewrite in terms of  $y_1$ ,  $y_2^*$ ,  $y_3^*$ , and  $y_4^*$ .

- ii. Use  $y_1$  to proxy  $h$ .
- iii. Use a cross-member IV for  $y_1$  in addition to  $y_j^*$ ,  $j = 2, 3, 4$  which gives our estimate of  $\nu_{45}$ .
- iv. Now form  $\tilde{y}_5 = y_5 - \nu_{45}y_4$ .
- v. With  $y_4$  excluded, we can use purged IV's on  $\tilde{y}_5$ , as before.

### 3.4 Comments

1. One needs to check the rank order conditions for identification (requires imposing an exclusion restriction).
2. Griliches and Chamberlain (IER, 1976) find a small ability bias -  $3^{rd}$  decimal point difference in schooling coefficient.

## 4. Twin Methods

Basic Principle: Monozygotic or MZ

(identical) twins are more similar than

Dizygotic or DZ (fraternal) twins. The key

assumption is that if environmental factors are

the same for both types of twins, then we can

estimate genetic components to outcomes.

### 4.1. Univariate Twin Model

Let  $y$  = observed phenotypic variable,

$x$  = unobserved genotype, and  $u$  = environ-

ment. Further, suppose that we can write our

model additively

$$y = x + u,$$

and assume independence of  $x$  and  $u$  so that

$$\sigma_y^2 = \sigma_x^2 + \sigma_u^2. \text{ Now suppose that we have data}$$

on another individual,  $y^* = x^* + u^*$ .

Then our phenotypic covariance is

$$Cov(y, y^*) = Cov(x, x^*) + Cov(u, u^*),$$

where we are imposing the assumption

$$Cov(x, u^*) = Cov(x^*, u) = 0. \text{ Defining}$$

standardized forms and some simplifying

notation, let

$$\tilde{y} = \frac{y}{\sigma_y}, \quad \tilde{x} = \frac{x}{\sigma_x}, \quad \tilde{u} = \frac{u}{\sigma_u},$$

$$h^2 = \frac{\sigma_x^2}{\sigma_y^2}, \rho^2 = \frac{\sigma_u^2}{\sigma_y^2}.$$

Thus,  $\tilde{y}\sigma_y = \tilde{x}\sigma_x + \tilde{u}\sigma_u$  which implies

$\tilde{y} = h\tilde{x} + \rho\tilde{u}$ . We can also derive the identity

$$h^2 + \rho^2 = \frac{\sigma_x^2}{\sigma_y^2} + \frac{\sigma_u^2}{\sigma_y^2} = 1$$

where the last step follows from our assumption of independence. Now we wish to consider the correlation between observed phenotypes of our two individuals:

$$\begin{aligned} C &= \text{Corr}(y, y^*) \\ &= \text{Corr}(h\tilde{x} + p\tilde{u}, h\tilde{x}^* + \rho\tilde{u}^*) \\ &= h^2 \frac{\text{Cov}(\tilde{x}, \tilde{x}^*)}{\text{Var}(\tilde{x})} + \rho^2 \frac{\text{Cov}(\tilde{u}, \tilde{u}^*)}{\text{Var}(\tilde{u})} \end{aligned}$$

$$= h^2g + \rho^2\nu,$$

say, with  $g$  and  $\nu$  defined as above. We assume that  $g_{MZ} = 1$  and that  $g_{DZ} < 1$ . That is, the genotypic variable is perfectly correlated among identical twins, but less than perfectly correlated among fraternal twins. Replacing this result into the above produces

$$C_{MZ} = h^2 + \nu_{MZ}\rho^2, C_{DZ} = h^2g_{DZ} + \nu_{DZ}\rho^2.$$

Therefore,

$$C_{MZ} - C_{DZ} = (1 - g_{DZ})h^2 + (\nu_{MZ} -$$

$$\begin{aligned} & \nu_{DZ})\rho^2 \\ & = (1 - g_{DZ})h^2 + (\nu_{MZ} - \nu_{DZ})(1 - h^2), \end{aligned}$$

where the last equality follows from our

established identity. Solving for  $h^2$ , we find

$$h^2 = \frac{(C_{MZ} - C_{DZ}) - (\nu_{MZ} - \nu_{DZ})}{(1 - g_{DZ}) - (\nu_{MZ} - \nu_{DZ})}.$$

The only known in the right hand side of the above equality is the expression  $(C_{MZ} - C_{DZ})$ , which is simply the correlation coefficient of the observed phenotypic variable. The remaining two expressions,  $(1 - g_{DZ})$  and  $(\nu_{MZ} - \nu_{DZ})$  can not be

computed as they represent statistics on variables we don't observe. One could impose

$\nu_{MZ} = \nu_{DZ}$  so that

$$h^2 = \frac{C_{MZ} - C_{DZ}}{1 - g_{DZ}}.$$

The expression  $g_{DZ}$  is a measure of how closely the genetic variable is correlated across our two observations. One could then guess or estimate a value for this parameter to derive corresponding estimates of  $h^2$ , the ratio of how much variance in the phenotypic variable is explained by variance in the genetic

component. Other studies have attempted to include  $Cov(x, u) \neq 0$  but this presents an identification problem. A typical value of the estimable portion of the above,  $C_{MZ} - C_{DZ}$ , is commonly reported in the literature to be 0.2.

## Appendix

Equation 2.2

We can write

$$y_t = x^* \beta + (U_t - \epsilon_{1t} \beta_1 - \epsilon_{2t} \beta_2),$$

where

$$x^* = \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

and  $x_1^*, x_2^*$ , are  $T \times 1$ . So,

$$\begin{aligned}
\hat{\beta}^{OLS} &= (x^{*'} x^*)^{-1} (x^{*'} y) \\
&= \beta + (x^{*'} x^*)^{-1} (x^{*'} (U - \epsilon_1 \beta_1 - \epsilon_2 \beta_2)) \\
&= \beta + \left( \frac{(x^{*'} x^*)}{T} \right)^{-1} \\
&\quad \left( \left( \frac{x^{*'} U}{T} \right) - \left( \frac{x^{*'} \epsilon_1 \beta_1}{T} \right) - \left( \frac{x^{*'} \epsilon_2 \beta_2}{T} \right) \right) \\
&\rightarrow \beta + (E(x^{*'} x^*))^{-1} \\
&\quad (E(x^{*'} U) - E(x^{*'} \epsilon_1) \beta_1 - E(x^{*'} \epsilon_2) \beta_2) \\
&= \beta - \left[ \begin{array}{cc} E(x_1^{*'} x_1^*) & E(x_1^{*'} x_2^*) \\ E(x_2^{*'} x_1^*) & E(x_2^{*'} x_2^*) \end{array} \right]^{-1} \\
&\quad \left( E \begin{bmatrix} x_1^{*'} \epsilon_1 \\ x_2^{*'} \epsilon_1 \end{bmatrix} \beta_1 + E \begin{bmatrix} x_1^{*'} \epsilon_2 \\ x_2^{*'} \epsilon_2 \end{bmatrix} \beta_2 \right) \\
&= \beta - \left[ \begin{array}{cc} E(x_1^{*'} x_1^*) & E(x_1^{*'} x_2^*) \\ E(x_2^{*'} x_1^*) & E(x_2^{*'} x_2^*) \end{array} \right]^{-1} \\
&\quad \left[ \begin{array}{cc} E(x_1^{*'} \epsilon_1) & E(x_1^{*'} \epsilon_2) \\ E(x_2^{*'} \epsilon_1) & E(x_2^{*'} \epsilon_2) \end{array} \right] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
&= \left( I - \begin{bmatrix} E(x_1^* x_1^*) & E(x_1^* x_2^*) \\ E(x_2^* x_1^*) & E(x_2^* x_2^*) \end{bmatrix}^{-1} \right) \\
&\quad \begin{bmatrix} E(x_1^* \epsilon_1) & E(x_1^* \epsilon_2) \\ E(x_2^* \epsilon_1) & E(x_2^* \epsilon_2) \end{bmatrix} \\
&\quad \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \\
&= \left( I - (\Sigma_{x^*})^{-1} \begin{bmatrix} E\left(\left(\begin{smallmatrix} \epsilon_1' + x_1' \\ \epsilon_2' + x_2' \end{smallmatrix}\right) \epsilon_1\right) & E\left(\left(\begin{smallmatrix} \epsilon_1' + x_1' \\ \epsilon_2' + x_2' \end{smallmatrix}\right) \epsilon_2\right) \\ E\left(\left(\begin{smallmatrix} \epsilon_1' + x_1' \\ \epsilon_2' + x_2' \end{smallmatrix}\right) \epsilon_1\right) & E\left(\left(\begin{smallmatrix} \epsilon_1' + x_1' \\ \epsilon_2' + x_2' \end{smallmatrix}\right) \epsilon_2\right) \end{bmatrix} \right) \\
&\quad \left( I - (\Sigma_{x^*})^{-1} (\Sigma_\epsilon) \right) \beta,
\end{aligned}$$

where the second-to-last step follows from the independence of  $x$  and  $\epsilon$ . This type of argument is also used to derive the probability limit of the  $\beta'$ s in section 1.