The returns to schooling, ability bias, and regression

Jörn-Steffen Pischke

LSE

October 4, 2016

Counterfactual outcomes

- Scholing for individual i is described by a binary random variable,

$$D_i = \{0, 1\}$$

  say $D_i = 1$ denotes completing high school, and $D_i = 0$ denotes dropping out.
- The outcome of interest, log earnings is denoted by y_i.
- Potential outcomes: What would have happened to someone who completes if they had dropped out and vice versa. Hence, for everybody there are two

$$\textit{potential outcomes} = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}.$$

Potential and observed outcomes

- The effect of completing school for individual i is $Y_{1i} - Y_{0i}$.
- The observed outcome, Y_i, can be written in terms of potential outcomes as

$$
\begin{aligned}
Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})D_i.
\end{aligned}
$$

- We only observe either Y_{1i} or Y_{0i} for a single individual.

The selection problem

We can write the observed difference in earnings for the treated and untreated as the sum of two terms:

$$\underbrace{E\left[Y_i|D_i=1\right]-E[Y_i|D_i=0]}_{\text{Observed diff in earnings cond. on schooling}} = E\left[Y_{1i}|D_i=1\right]-E[Y_{0i}|D_i=0]$$

$$= \underbrace{E\left[Y_{1i}|D_i=1\right]-E[Y_{0i}|D_i=1]}_{\text{average treatment effect on the treated}}$$

$$+\underbrace{E\left[Y_{0i}|D_i=1\right]-E\left[Y_{0i}|D_i=0\right]}_{\text{ability or selection bias}}$$

Selection on observables

The key to regression is that conditional on observable variables, e.g. ability, counterfactual outcomes are mean independent of schooling:

$$E\left[Y_{0i}|A_i, D_i\right] = E\left[Y_{0i}|A_i\right] \tag{1}$$

Then:

$$\begin{aligned}
&\underbrace{E\left[Y_i|A_i, D_i = 1\right] - E\left[Y_i|A_i, D_i = 0\right]}_{\text{Observed difference in average log earnings}} \\
= \quad & E\left[Y_{1i}|A_i, D_i = 1\right] - E\left[Y_{0i}|A_i, D_i = 0\right] \\
= \quad & E\left[Y_{1i}|A_i, D_i = 1\right] - \underbrace{E\left[Y_{0i}|A_i, D_i = 1\right]}_{\text{if (1) holds get to switch } D_i} \\
= \quad & \underbrace{E\left[Y_{1i} - Y_{0i}|A_i, D_i = 1\right]}_{\text{average treatment effect on the treated in groups defined by } A_i}
\end{aligned}$$

Selection on observables leads to matching

We have just estimated group specific treatment effects defined by the variable A_i:

$$\begin{aligned} & E\left[Y_i | A_i = a, D_i = 1\right] - E\left[Y_i | A_i = a, D_i = 0\right] \\ = \; & E\left[Y_{1i} - Y_{0i} | A_i = a, D_i = 1\right] \equiv \rho_a \end{aligned}$$

Do this for all values the variable A_i takes on.
To get from these group conditional averages to the overall average use the rules for conditional probability

$$\begin{aligned} & E\left[Y_{1i} - Y_{0i} | D_i = 1\right] \\ = \; & \sum_a E\left[Y_{1i} - Y_{0i} | A_i = a, D_i = 1\right] \Pr\left(A_i = a | D_i = 1\right) \\ = \; & \sum_a \rho_a \Pr\left(A_i = a | D_i = 1\right) \end{aligned}$$

From matching to regression

How to get from this to a regression is easiest to see in a constant effects setting. So let $\rho_a = \rho$ and write

$$\begin{aligned} Y_{0i} &= \alpha + \eta_i \\ Y_{1i} &= Y_{0i} + \rho \end{aligned}$$

Using

$$\begin{aligned} Y_i &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i \\ &= \underbrace{E(Y_{0i})}_{\alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\rho} D_i + \underbrace{[Y_{0i} - E(Y_{0i})]}_{\eta_i} \\ &= \alpha + \rho D_i + \eta_i \end{aligned}$$

Still on the road to regression

$$Y_i = \alpha + \rho D_i + \eta_i \tag{2}$$

Using the CIA (1)

$$E[Y_{0i}|A_i, D_i] = E[Y_{0i}|A_i] \Leftrightarrow E[\eta_i|A_i, D_i] = E[\eta_i|A_i]$$

we see that η_i is a function of A_i, which is related to D_i. So (2) is not a regression yet. Approximate $E[\eta_i|A_i]$ by a linear function

$$\begin{aligned} E[\eta_i|A_i] &= \gamma A_i \\ \eta_i &= \gamma A_i + e_i \qquad E[e_i|A_i] = 0 \end{aligned}$$

Substitute in (2):

$$Y_i = \alpha + \rho D_i + \gamma A_i + e_i$$

which is a regression (i.e. $E[e_i|A_i, D_i] = 0$).

Regression anatomy

A key tool in understanding multivariate regression the *regression anatomy formula* (or Frisch-Waugh-Lowell theorem). Consider a generic bivariate regression:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

Then

$$\beta_1 = \frac{Cov(Y_{i,} \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})}$$

where $\widetilde{X}_{1i}$ is the residual from a regression of X_{1i} on X_{2i}:

$$X_{1i} = \pi_0 + \pi_1 X_{2i} + \widetilde{X}_{1i}.$$

The regression anatomy formula derived

Substitute the expression for the long regression into

$$\begin{aligned}
\frac{Cov(Y_i, \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})} &= \frac{Cov(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i, \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})} \\
&= \frac{\beta_1 \overbrace{Cov(X_{1i}, \widetilde{X}_{1i})}^{=Var(\widetilde{X}_{1i})} + \beta_2 \overbrace{Cov(X_{2i}, \widetilde{X}_{1i})}^{=0} + \overbrace{Cov(e_i, \widetilde{X}_{1i})}^{=0}}{Var(\widetilde{X}_{1i})} \\
&= \beta_1 \frac{Var(\widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})} = \beta_1.
\end{aligned}$$

In words: the coefficient from a bivariate regression of Y_i on $\widetilde{X}_{1i}$ is numerically the same as the long regression coefficient on X_{1i} from a regression of Y_i on X_{1i} and X_{2i}.

Alternative versions of regression anatomy

- Regression anatomy can also be written as

$$\beta_1 = \frac{Cov(Y_i, \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})} = \frac{Cov(\widetilde{Y}_i, \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})}$$

  i.e. also partialling out X_{2i} from the dependent variable.

- It is not enough to partial out the covariate from the dependent variable alone

$$\frac{Cov(\widetilde{Y}_i, X_{1i})}{Var(X_{1i})} = \frac{Cov(\widetilde{Y}_i, \widetilde{X}_{1i})}{Var(\widetilde{X}_{1i})} \frac{Var(\widetilde{X}_{1i})}{Var(X_{1i})} \neq \beta_1$$

  (you *can* partial out X_{2i} from Y_i but you *must* partial it out from X_{1i}).

- Regression anatomy works for multiple covariates just as well.

Ability bias and the returns to schooling

We would like to run the *long regression*

$$Y_i = \alpha + \rho S_i + \gamma A_i + e_i$$

where Y_i is log earnings, S_i is schooling and A_i is ability. If we don't have a measure of ability we can only run the *short regression*

$$Y_i = \alpha_s + \rho_s S_i + e_i^s.$$

What do we get?

The omitted variables bias formula

- The relationship between the long and short regression coefficients is given by the *omitted variables bias (OVB)* formula

$$\rho_s = \frac{Cov(Y_i, S_i)}{Var(S_i)} = \rho + \gamma\delta_{AS}$$

where

$$\delta_{AS} = \frac{Cov(A_i, S_i)}{Var(S_i)}$$

is the regeression coefficient from a regression of A_i (the omitted variable) on S_i (the included variable).

  - Exercise: Derive the OVB formula (works just like for regression anatonmy).

- The OVB formula is a mechanical relationship between two regressions: it holds regardless of the causal interpretation of any of the coeffcients.

Griliches (1977) regressions

- The conventional wisdom is $Cov(A_{i,} S_i) > 0$, so returns to schooling estimates will be biased up.
- Short regression estimates using the NLS

$$Y_i = \text{const} + \underset{(0.003)}{0.068}\, S_i + \text{experience}$$

- Long regression estimates

$$Y_i = \text{const} + \underset{(0.003)}{0.059}\, S_i + \underset{(0.0005)}{0.0028}\ \text{IQ}_i + \text{experience}$$

- The results are consistent with the conventional wisdom.

Classical measurement error

- Measurement error leads to bias. Many economic variables are mismeasured.
- Look at a generic example and start with a simple bivariate regression

$$Y_i = \alpha + \beta X_i^* + e_i.$$

We don't observe X_i^* but X_i

$$X_i = X_i^* + m_i$$

where

$$\begin{aligned} Cov(X_i^*, m_i) &= 0 \\ Cov(e_i, m_i) &= 0. \end{aligned}$$

This is called *classical measurement error*.

Attenuation from classical measurement error

The bivariate regression coefficient we estimate is

$$\begin{aligned}\widehat{\beta} &= \frac{Cov(Y_i, X_i)}{Var(X_i)} \\ &= \frac{Cov(\alpha + \beta X_i^* + e_i, X_i^* + m_i)}{Var(X_i^* + m_i)} \\ &= \beta \frac{Var(X_i^*)}{Var(X_i^*) + Var(m_i)} = \beta\lambda.\end{aligned}$$

We see that β is biased towards zero by an attenuation factor

$$\lambda = \frac{Var(X_i^*)}{Var(X_i^*) + Var(m_i)}$$

which is the variance in the "signal" divided by the variance in the "signal plus noise."

How measurement error works

Measurement error in the returns to schooling

- Think of Y_i as log earnings, and X_i as schooling. Ignore age or experience for the moment.
- Ashenfelter and Krueger (1994) find $\lambda = 0.9$ for schooling.
- This means if the true return to schooling is 0.1, we would expect an estimate of 0.09.

Measurement error with two regressors

- The bivariate regression is not particularly interesting, since we typically want to use regression to control for other factors.
- Consider

$$Y_i = \alpha + \beta_1 X_{1i}^* + \beta_2 X_{2i} + e_i.$$

and *only* X_{1i}^* is subject to classical measurement error, i.e. $Cov(X_{1i}^*, m_i) = Cov(X_{2i}, m_i) = 0$.

- Starting from

$$X_{1i} = X_{1i}^* + m_i$$

and using the classical measurement error assumptions we have

$$\widetilde{X}_{1i} = \widetilde{X}_{1i}^* + m_i.$$

Measurement error with two regressors

The estimator of β_1 is now (using regression anatomy):

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{Cov\left(Y_i, \widetilde{X}_{1i}\right)}{Var\left(\widetilde{X}_{1i}\right)} \\
&= \frac{Cov\left(\alpha + \beta_1 X_{1i}^* + \beta_2 X_{2i} + e_i, \widetilde{X}_{1i}^* + m_i\right)}{Var\left(\widetilde{X}_{1i}\right)} \\
&= \beta_1 \frac{Var\left(\widetilde{X}_{1i}^*\right)}{Var\left(\widetilde{X}_{1i}^*\right) + Var\left(m_i\right)} = \beta_1 \widetilde{\lambda}
\end{aligned}$$

It is easy to see that $\widetilde{\lambda} \leq \lambda$ whenever X_{1i}^* and X_{2i} are correlated.

Fact

Adding correlated regressors makes attenuation bias from measurement error worse.

Comparing the short and long regression

- The short regression (on just X_{1i}) coefficient is

$$\widehat{\beta}_{1,short} = \lambda\beta_1 + \beta_2\delta_{X_2X_1} = \lambda\left(\beta_1 + \beta_2\delta_{X_2X_1^*}\right)$$

where the estimate of β_1 is biased both because of attenuation due to measurement error, and because of omitted variables bias (the part $\beta_2\delta_{X_2X_1}$ where $\delta_{X_2X_1}$ is the coefficient from a regression of X_{2i} on X_{1i}).

- The coefficient from the long regression is

$$\widehat{\beta}_{1,long} = \beta_1\widetilde{\lambda}$$

and since

$$\widetilde{\lambda} < \lambda$$

but

$$\widehat{\beta}_{1,short} \lesseqgtr \widehat{\beta}_{1,long}.$$

Comparing the short and long regression

- Notice that it is more difficult to compare the bias from the short regression and the long regression now.
- $\widetilde{\lambda} < \lambda$ implies that the attenuation bias goes up when another regressor is entered which is correlated with X_{1i}.
- There is less attenuation in the short regression but there is also OVB now. Not clear what the net effect is.

Measurement error in the control

- What about the coefficient β_2? Even when there is no measurement error in X_{2i}, the estimate of β_2 will be biased:

$$\widehat{\beta}_2 = \beta_1 \delta_{X_2 X_1^*} \left(1 - \widetilde{\lambda}\right) + \beta_2.$$

- Note that the bias will be larger the larger
  - the measurement error
  - the correlation between X_{1i}^* and X_{2i}
- The intuition is that
  - β_1 is attenuated, and hence does not reflect the full effect of X_{1i}^*
  - β_2 will capture part of the effect of X_{1i}^* through the correlation with X_{2i}

Measurement error in the returns to schooling controlling for ability

We want to run the regression

$$y_i = \alpha + \rho S_i^* + \gamma A_i + e_i$$

where S_i^* is schooling and A_i is ability. Suppose we only have a mismeasured version of schooling, S_i (so S_i takes on the role of X_{1i} before, and A_i takes on the role of X_{2i}). Then the short regression will give

$$\widehat{\rho}_{short} = \lambda\rho + \gamma\delta_{AS}$$

and the long regression

$$\widehat{\rho}_{long} = \widetilde{\lambda}\rho$$

If ability bias is upwards ($\delta_{AS} > 0$) it is not possible to say a priori which estimate will be closer to ρ.

Putting some numbers on the Griliches example

Pick some numbers for the regression

$$y_i = 0.1S_i^* + 0.01A_i + e_i$$

and set

$$\begin{aligned} \lambda &= 0.9 \\ \sigma_S &= 3, \sigma_A = 15, \sigma_{AS} = 22.5. \end{aligned}$$

Then

$$\delta_{AS} = \frac{\sigma_{AS}}{\sigma_S^2} = \frac{22.5}{9} = 2.5$$

and

$$\widehat{\rho}_{short} = \lambda\rho + \gamma\delta_{AS} = 0.9 \times 0.1 + 0.01 \times 2.5 = 0.115$$

What about the long regression?

We first need

$$\widetilde{\lambda} = \frac{Var\left(\widetilde{S}_i^*\right)}{Var\left(\widetilde{S}_i^*\right) + Var\left(m_i\right)} = \frac{\lambda - R_{AS}^2}{1 - R_{AS}^2}$$

which is

$$R_{AS}^2 = \left(\frac{\sigma_{AS}}{\sigma_S \sigma_A}\right)^2 = \left(\frac{22.5}{45}\right)^2 = 0.25$$

$$\widetilde{\lambda} = \frac{0.9 - 0.25}{1 - 0.25} = 0.867.$$

Then the long regression coefficient is

$$\widehat{\rho}_{long} = \widetilde{\lambda}\rho = 0.867 \times 0.1 = 0.087$$

so the short regression coefficient is too large and the long regression coefficient is too small.

Measurement error in ability

Now suppose years of schooling is measured perfectly but instead of A_i^* we only have mismeasured ability A_i (so S_i takes on the role of X_{2i} before, and A_i takes on the role of X_{1i}). Then

$$\widehat{\rho} = \gamma\delta_{AS}\left(1 - \widetilde{\lambda}\right) + \rho.$$

If ability bias is upwards ($\delta_{AS} > 0$) then the returns to schooling will be biased up but by less than in the short regression. Controlling for A_i is better than controlling for nothing, but not as good as controlling for true ability A_i^*.

Instrumental variables solve the measurement error problem

- Suppose you have an instrument Z_i, correlated with the signal X_i^* and uncorrelated with the error m_i.
- In the bivariate regression you get

$$\widehat{\beta}_{IV} = \frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} = \frac{Cov(\alpha + \beta X_i^* + e_i, Z_i)}{Cov(X_i^* + m_i, Z_i)} = \frac{\beta Cov(X_i^*, Z_i)}{Cov(X_i^*, Z_i)} = \beta$$

- In the multivariate regression you get for similar reasons:

$$\begin{aligned} \widehat{\beta}_{1,IV} &= \beta_1 \\ \widehat{\beta}_{2,IV} &= \beta_2 \end{aligned}$$

(notice that only the mismeasured X_{1i} is instrumented)

More Griliches (1977) regressions

Recall the Griliches long regression estimates

$$y_i = \text{const} + \underset{(0.003)}{0.059}\, S_i + \underset{(0.0005)}{0.0028}\ \text{IQ}_i + \text{experience}$$

Instrumenting IQ with results from the Knowledge of the World of Work test he gets

$$y_i = \text{const} + \underset{(0.004)}{0.052}\, S_i + \underset{(0.0009)}{0.0051}\ \text{IQ}_i + \text{experience}$$

- This is still consistent with two stories:

More Griliches (1977) regressions

Recall the Griliches long regression estimates

$$y_i = \text{const} + \underset{(0.003)}{0.059}\, S_i + \underset{(0.0005)}{0.0028}\ \text{IQ}_i + \text{experience}$$

Instrumenting IQ with results from the Knowledge of the World of Work test he gets

$$y_i = \text{const} + \underset{(0.004)}{0.052}\, S_i + \underset{(0.0009)}{0.0051}\ \text{IQ}_i + \text{experience}$$

- This is still consistent with two stories:
- there is upward ability bias in the bivariate return to schooling and measurement error in IQ but not schooling (in this case the estimate of 0.052 is correct)

More Griliches (1977) regressions

Recall the Griliches long regression estimates

$$y_i = \text{const} + \underset{(0.003)}{0.059} S_i + \underset{(0.0005)}{0.0028} \text{ IQ}_i + \text{experience}$$

Instrumenting IQ with results from the Knowledge of the World of Work test he gets

$$y_i = \text{const} + \underset{(0.004)}{0.052} S_i + \underset{(0.0009)}{0.0051} \text{ IQ}_i + \text{experience}$$

- This is still consistent with two stories:
- there is upward ability bias in the bivariate return to schooling and measurement error in IQ but not schooling (in this case the estimate of 0.052 is correct)
- there is measurement error in both schooling and ability (in this case the true coefficient could be anything).

What to control for?

In the quest for identifying causal effects, which variables belong on the right hand side of a regression equation?

- Yes: Variables determining the treatment and correlated with the outcome (e.g. ability).
  - in general these variables will be fixed characteristics or pre-determined by the time of treatment (e.g. schooling)
  - but remember the warnings on measurement error just discussed: the kitchen sink is no panacea
- Yes: Variables uncorrelated with the treatment but correlated with the outcome
  - these variables may help reducing standard errors
- No: Variables which are outcomes of the treatment itself. These are *bad controls*.

Bad control

Some researchers regressing earnings on schooling (and experience) include controls for occupation. Does this make sense?

- Clearly we can think of schooling affecting the access to higher level occupations, e.g. you need a Ph.D. to become a college professor. This gives rise to a two equation system

$$\begin{aligned} Y_i &= \alpha + \rho S_i + \gamma O_i + e_i \\ O_i &= \lambda_0 + \lambda_1 S_i + u_i \end{aligned}$$

  You could think about these as a simultaneous equations system. Occupation O_i is an endogenous variable. As a result, you could not necessarily estimate the first equation by OLS.

- Occupation is a *bad control*.

Bad control example 1

	occupation		*wage*		*Observed data*	
	O_0	O_1	W_0	W_1	$S = 0$	$S = 1$
Type 1	B	B	600	600	(B, 625)	(B, 600)
Type 2	B	W	650	700		(W, 700)
Type 3	W	W	700	700	(W, 700)	

Bad control example 2

	occupation		*wage*		*Observed data*	
	O_0	O_1	W_0	W_1	$S = 0$	$S = 1$
Type 1	B	B	600	625	(B, 625)	(B, 625)
Type 2	B	W	650	700		(W, 725)
Type 3	W	W	725	750	(W, 725)	

Proxy control

Sometimes we control for a variable in the best of intentions. Suppose our regression of schooling on earnings

$$Y_i = \alpha + \rho S_i + \gamma A_i + e_i$$

has a causal interpretation conditional on ability.
Instead of ability we only have a test score taken at age 18, call it A_{li} for late ability. The problem is that schooling will already have influenced the late ability (some students will have dropped out by 18). Suppose

$$A_{li} = \pi_0 + \pi_1 S_i + \pi_2 A_i$$

i.e. that age 18 test scores are influenced by both schooling and true ability.

What do we get with proxy control?

Substituting for A_i in our regression above we get

$$y_i = \left(\alpha - \gamma \frac{\pi_0}{\pi_2}\right) + \left(\rho - \gamma \frac{\pi_1}{\pi_2}\right) S_i + \frac{\gamma}{\pi_2} A_{Ii} + e_i .$$

If

$$\begin{aligned} \rho &> 0 \\ \gamma &> 0 \\ \pi_1 &> 0, \pi_2 > 0 \end{aligned}$$

then

$$\left(\rho - \gamma \frac{\pi_1}{\pi_2}\right) < \rho$$

and we will estimate a return to schooling that is too small.

Are our results robust?

Suppose you have an identification strategy, say

$$Y_i = \alpha + \rho S_i + \gamma_1 A_i + e_i.$$

You have another variable, X_i, which could potentially be an additional confounder. What do you do with it?

- $$Y_i = \alpha_l + \rho_l S_i + \gamma_{1l} A_i + \gamma_2 X_i + e_i^l$$

  Is $\rho_l = \rho$?

- $$X_i = \delta_0 + \delta S_i + \delta_1 A_i + u_i$$

  Is $\delta = 0$?

A tale of two tests

$$\begin{aligned} Y_i &= \alpha + \rho S_i + \gamma_1 A_i + e_i \\ Y_i &= \alpha_l + \rho_l S_i + \gamma_{1l} A_i + \gamma_2 X_i + e_i^l \\ X_i &= \delta_0 + \delta S_i + \delta_1 A_i + u_i \end{aligned}$$

- The test $\rho_l = \rho$ is a coefficient comparison test. Formally a generalized Hausman test.
- The test $\delta = 0$ is a balancing test.

How are they related? Through the OVB formula:

$$\rho - \rho_l = \gamma_2 \delta.$$

Under the maintained assumption $\gamma_2 \neq 0$, the two tests test the same null hypothesis $\delta = 0$.

Measurement error in controls again

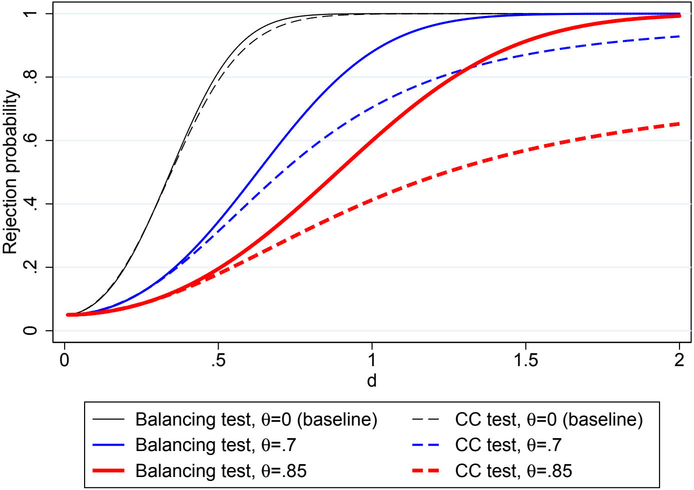
Remember

$$\begin{aligned} \widehat{\rho}_l &= \gamma_2 \delta \left(1 - \widetilde{\lambda}\right) + \rho \\ \widehat{\gamma}_2 &= \gamma_2 \widetilde{\lambda}. \end{aligned}$$

- Classical measurement error in X_i biases $\widehat{\rho}_l$ towards ρ, and $\widehat{\gamma}_2$ towards 0.
  - $\widehat{\rho}_l$ closer to ρ: power of the coefficient comparison test directly affected.
  - Variation from measurement error ends up in e_i^l, raising standard errors and reducing power
- There is no bias in the estimate of δ
  - Variation from measurement error ends up in u_i, raising standard errors and reducing power

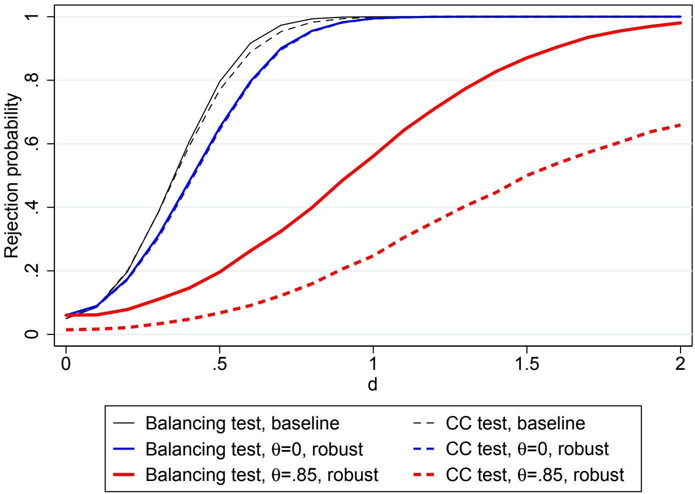
Theoretical Power Functions

Homoskedastic, plain standard errors



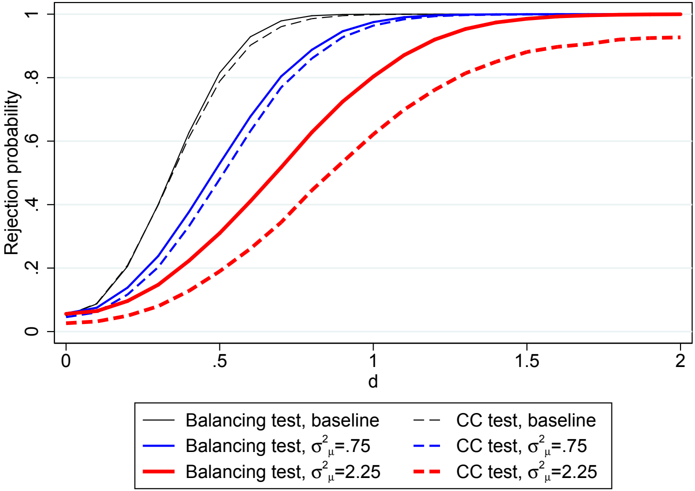
Simulated Rejection Rates

Heteroskedasticity, robust standard errors



Simulated Rejection Rates

Mean reverting measurement error, robust standard errors



Baseline Regressions for Returns to Schooling

	Log hourly earnings			Mother's years of education	Library card at age 14
	(1)	(2)	(3)	(6)	(7)
Years of education	0.0751 (0.0040)	0.0728 (0.0042)	0.0735 (0.0040)	0.3946 (0.0300)	0.0371 (0.0040)
Mother's years of education		0.0059 (0.0029)			
Library card at age 14			0.0428 (0.0183)		
p-values					
Coefficient comparison test		0.045	0.023		
Balancing test				0.000	0.000

Returns to Schooling controlling for KWW score

	Log hourly earnings			Mother's years of education	Library card at age 14
	(1)	(2)	(3)	(6)	(7)
Years of education	0.0609 (0.0059)	0.0596 (0.0060)	0.0608 (0.0059)	0.2500 (0.0422)	0.0133 (0.0059)
KWW score	0.0070 (0.0015)	0.0068 (0.0016)	0.0069 (0.0016)	0.0410 (0.0107)	0.0076 (0.0016)
Mother's years of education		0.0053 (0.0037)			
Library card at age 14			0.0097 (0.0215)		
Body height in inches					
p-values					
Coefficient comparison test		0.163	0.652		
Balancing test				0.000	0.025

Returns to Schooling instrumenting the KWW score

	Log hourly earnings			Mother's years of education	Library card at age 14
	(1)	(2)	(3)	(6)	(7)
Years of education	0.0340	0.0339	0.0342	0.0234	0.0168
	(0.0139)	(0.0139)	(0.0138)	(0.0952)	(0.0134)
KWW score instrumented by IQ	0.0194	0.0195	0.0200	0.1496	0.0060
	(0.0063)	(0.0063)	(0.0063)	(0.0422)	(0.0060)
Mother's years of education		0.0028			
		(0.0039)			
Library card at age 14			-0.0130		
			(0.0245)		
Body height in inches					
p-values					
Coefficient comparison test		0.818	0.635		
Balancing test				0.806	0.212