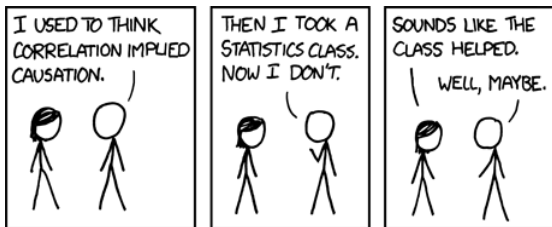


## Advanced Applied Econometrics

Teacher: Felix Weinhardt



## Outline: Selected OLS topics

- Regression anatomy theorem
- Omitted variable bias
- Interpreting coefficient movements
- Heterogeneous effects

Slides are partly based on material provided by Scott Cunningham.

- Regression anatomy theorem

### Theorem 3.1.7: Regression anatomy theorem

- Regression anatomy theorem is maybe more intuitive with an example and some data visualization. It concerns multiple linear regression.
- Can we estimate the causal effect of family size on labor supply by regressing labor supply (`workforpay`) on family size (`numkids`)?

$$workforpay_i = \beta_0 + \beta_1 numkids_i + u_i$$

```
. regress workforpay numkids
```

where the first line is the causal / econometric model, and the second line is the regression command in STATA

- If family size is random, then number of kids is uncorrelated with the unobserved error term, which means we can interpret  $\hat{\beta}_1$  as the causal effect.
  - Example: if Melissa has no children in reality (i.e., `numkids`= 0) and we wanted to know what the effect on labor supply will be if we surgically manipulated her family size (i.e., `numkids` = 1) then  $\hat{\beta}_1$  would be our answer
  - Visual: Even better, we could just plot the regression coefficient in a scatter plot showing all  $i$  (`workforpay`, `numkids`) pairs and the slope coefficient would be the best fit of the data through these points, as well as tell us the average causal effect of family size on labor supply
- But how do we interpret  $\hat{\beta}_1$  if `numkids` is non-random?

- Assume that family size is random once we condition on race, age, marital status and employment. Then the model is:

$$\text{Workforpay}_i = \beta_0 + \beta_1 \text{Numkids}_i + \gamma_1 \text{White}_i + \gamma_2 \text{Married}_i + \gamma_3 \text{Age}_i + \gamma_4 \text{Employed}_i + u_i$$

- If we want to estimate average causal effect of family size on labor supply, we will need two things:
  - ① a data set with all 6 variables;
  - ② numkids must be randomly assigned conditional on the other 4 variables
- Now how do we interpret  $\hat{\beta}_1$ ? And can we visualize  $\hat{\beta}_1$  when there's multiple dimensions to the data? Yes, using the regression anatomy theorem, we can.

### Theorem 3.1.7: Regression Anatomy Theorem

Assume your main multiple regression model of interest:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \cdots + \beta_K x_{Ki} + e_i$$

and an auxiliary regression in which the variable  $x_{1i}$  is regressed on all the remaining independent variables

$$x_{1i} = \gamma_0 + \gamma_{k-1} x_{k-1i} + \gamma_{k+1} x_{k+1i} + \cdots + \gamma_K x_{Ki} + f_i$$

and  $\tilde{x}_{1i} = x_{1i} - \hat{x}_{1i}$  being the residual from the auxiliary regression. The parameter  $\beta_1$  can be rewritten as:

$$\beta_1 = \frac{\text{Cov}(y_i, \tilde{x}_{1i})}{\text{Var}(\tilde{x}_{1i})}$$

In words: The regression anatomy theorem is about interpretation. It says that  $\hat{\beta}_1$  is simply a scaled covariance with the  $\tilde{x}_1$  residual used instead of the actual data  $x$ .

I think a more detailed proof could be helpful, so I'm leaving it in the slides for now.

## Regression Anatomy Proof

To prove the theorem, note  $E[\tilde{x}_{ki}] = E[x_{ki}] - E[\hat{x}_{ki}] = E[f_i]$ , and plug  $y_i$  and residual  $\tilde{x}_{ki}$  from  $x_{ki}$  auxiliary regression into the covariance  $\text{cov}(y_i, \tilde{x}_{ki})$

$$\begin{aligned}\beta_k &= \frac{\text{cov}(y_i, \tilde{x}_{ki})}{\text{var}(\tilde{x}_{ki})} \\ &= \frac{\text{cov}(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \cdots + \beta_K x_{Ki} + e_i, \tilde{x}_{ki})}{\text{var}(\tilde{x}_{ki})} \\ &= \frac{\text{cov}(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \cdots + \beta_K x_{Ki} + e_i, f_i)}{\text{var}(f_i)}\end{aligned}$$

- 1 Since by construction  $E[f_i] = 0$ , it follows that the term  $\beta_0 E[f_i] = 0$ .
- 2 Since  $f_i$  is a linear combination of all the independent variables with the exception of  $x_{ki}$ , it must be that

$$\beta_1 E[f_i x_{1i}] = \cdots = \beta_{k-1} E[f_i x_{k-1i}] = \beta_{k+1} E[f_i x_{k+1i}] = \cdots = \beta_K E[f_i x_{Ki}] = 0$$



- ③ Consider now the term  $E[e_i f_i]$ . This can be written as:

$$\begin{aligned}
 E[e_i f_i] &= E[e_i f_i] \\
 &= E[e_i \tilde{x}_{ki}] \\
 &= E[e_i (x_{ki} - \hat{x}_{ki})] \\
 &= E[e_i x_{ki}] - E[e_i \tilde{x}_{ki}]
 \end{aligned}$$

Since  $e_i$  is uncorrelated with any independent variable, it is also uncorrelated with  $x_{ki}$ : accordingly, we have  $E[e_i x_{ki}] = 0$ . With regard to the second term of the subtraction, substituting the predicted value from the  $x_{ki}$  auxiliary regression, we get

$$E[e_i \tilde{x}_{ki}] = E[e_i (\hat{\gamma}_0 + \hat{\gamma}_1 x_{1i} + \cdots + \hat{\gamma}_{k-1} x_{k-1i} + \hat{\gamma}_{k+1} x_{k+1i} + \cdots + \hat{\gamma}_K x_{Ki})]$$

Once again, since  $e_i$  is uncorrelated with any independent variable, the expected value of the terms is equal to zero. Then, it follows  $E[e_i f_i] = 0$ .

- The only remaining term is  $E[\beta_k x_{ki} f_i]$  which equals  $E[\beta_k x_{ki} \tilde{x}_{ki}]$  since  $f_i = \tilde{x}_{ki}$ . The term  $x_{ki}$  can be substituted using a rewriting of the auxiliary regression model,  $x_{ki}$ , such that

$$x_{ki} = E[x_{ki}|X_{-k}] + \tilde{x}_{ki}$$

This gives

$$\begin{aligned} E[\beta_k x_{ki} \tilde{x}_{ki}] &= E[\beta_k E[\tilde{x}_{ki}(E[x_{ki}|X_{-k}] + \tilde{x}_{ki})]] \\ &= \beta_k E[\tilde{x}_{ki}(E[x_{ki}|X_{-k}] + \tilde{x}_{ki})] \\ &= \beta_k \{E[\tilde{x}_{ki}^2] + E[(E[x_{ki}|X_{-k}]\tilde{x}_{ki})]\} \\ &= \beta_k \text{var}(\tilde{x}_{ki}) \end{aligned}$$

which follows directly from the orthogonality between  $E[x_{ki}|X_{-k}]$  and  $\tilde{x}_{ki}$ . From previous derivations we finally get

$$\text{cov}(y_i, \tilde{x}_{ki}) = \beta_k \text{var}(\tilde{x}_{ki})$$

which completes the proof. □

## STATA command: reganat (i.e., regression anatomy)

```
. ssc install reganat, replace
. sysuse auto
. regress price length weight headroom mpg
. reganat price length weight headroom mpg, dis(weight length) biline
```

```
. regress price length weight headroom mpg
```

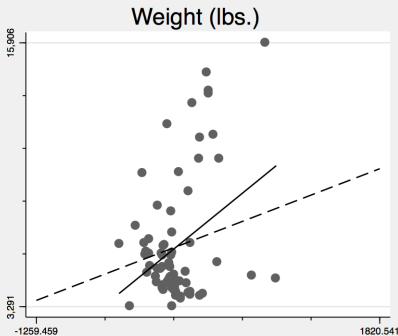
Source	SS	df	MS	Number of obs = 74		
Model	236190226	4	59047556.6	F( 4, 69) = 10.21		
Residual	398875170	69	5780799.56	Prob > F = 0.0000		
Total	635065396	73	8699525.97	R-squared = 0.3719		
				Adj R-squared = 0.3355		
				Root MSE = 2404.3		

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
length	-94.49651	40.39563	-2.34	0.022	-175.0836	-13.90944
weight	4.335045	1.162745	3.73	0.000	2.015432	6.654657
headroom	-490.9667	388.4892	-1.26	0.211	-1265.981	284.048
mpg	-87.95838	83.5927	-1.05	0.296	-254.7213	78.80449
_cons	14177.58	5872.766	2.41	0.018	2461.735	25893.43

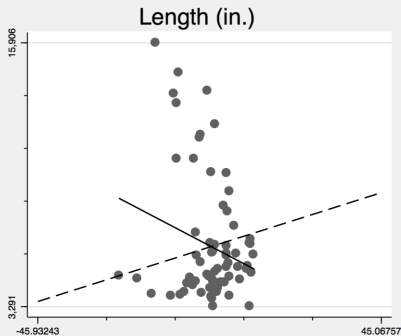
# Regression Anatomy

Dependent variable: Price



Multivariate slope: 4.335 (1.163)

Bivariate slope: 2.044 (0.377)



Multivariate slope: -94.497 (40.396)

Bivariate slope: 57.202 (14.080)

Covariates: Length (in.), Weight (lbs.), Headroom (in.), Mileage (mpg).

Regression lines: Solid = Multivariate, Dashed = Bivariate.

## Big picture

- 1 Regression provides the best linear predictor for the dependent variable in the same way that the CEF is the best unrestricted predictor of the dependent variable
- 2 If we prefer to think of approximating  $E(y_i|x_i)$  as opposed to predicting  $y_i$ , the regression CEF theorem tells us that even if the CEF is nonlinear, regression provides the best linear approximation to it.
- 3 Regression anatomy theorem helps us interpret a single slope coefficient in a multiple regression model by the aforementioned decomposition.

- Omitted variable bias formula

## Omitted Variable Bias

- A typical problem is when a key variable is omitted. Assume schooling causes earnings to rise:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + u_i$$

$Y_i$  = log of earnings

$S_i$  = schooling measured in years

$A_i$  = individual ability

- Typically the econometrician cannot observe  $A_i$ ; for instance, the Current Population Survey doesn't present adult respondents' family background, intelligence, or motivation.

- What are the consequences of leaving ability out of the regression?  
Suppose you estimated this short regression instead:

$$Y_i = \beta_0 + \beta_1 S_i + \eta_i$$

where  $\eta_i = \beta_2 A_i + u_i$ ;  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  are population regression coefficients;  $S_i$  is correlated with  $\eta_i$  through  $A_i$  only; and  $u_i$  is a regression residual uncorrelated with all regressors by definition.



## Derivation of Ability Bias

- Suppressing the  $i$  subscripts, the OLS estimator for  $\beta_1$  is:

$$\hat{\beta}_1 = \frac{\text{Cov}(Y, S)}{\text{Var}(S)} = \frac{E[YS] - E[Y]E[S]}{\text{Var}(S)}$$

- Plugging in the true model for  $Y$ , we get:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\text{Cov}[(\beta_0 + \beta_1 S + \beta_2 A + u), S]}{\text{Var}(S)} \\&= \frac{E[(\beta_0 S + \beta_1 S^2 + \beta_2 SA + uS)] - E(S)E[\beta_0 + \beta_1 S + \beta_2 A + u]}{\text{Var}(S)} \\&= \frac{\beta_1 E(S^2) - \beta_1 E(S)^2 + \beta_2 E(AS) - \beta_2 E(S)E(A) + E(uS) - E(S)E(u)}{\text{Var}(S)} \\&= \beta_1 + \beta_2 \frac{\text{Cov}(A, S)}{\text{Var}(S)}\end{aligned}$$

- If  $\beta_2 > 0$  and  $\text{Cov}(A, S) > 0$  the coefficient on schooling in the shortened regression (without controlling for  $A$ ) would be upward biased

## Summary

- When  $Cov(A, S) > 0$  then ability and schooling are correlated.
- When ability is unobserved, then not even multiple regression will identify the causal effect of schooling on wages.
- Here we see one of the main justifications for this class – what will we do when the treatment variable is endogenous? Because endogeneity means the causal effect has not been identified.

- Interpreting coefficient movements

## When can we be sure there is no OVB left in what we have?

- Imagine you start off from a situation where there are confounders, i.e.  $\text{Cov}(A, S) > 0$
- You can now add more controls and your  $\hat{\beta}_1$  estimate changes.
- This is an improvement –can we learn about how much is left to improve?

## Altonji, Elder, Taber: JPE 2005 idea

- Consider the following simple OLS model:

$$Y_i = \beta_0 + \beta_1 D_i + X_i \beta + \epsilon_i$$

- The OLS assumption is:

$$\text{Cov}(D, \epsilon) = 0$$

- This assumption implies that the error term  $\epsilon_i$  is uncorrelated with the independent variable  $D_i$ .

- Assume equality of selection on unobservables and observables:

$$\frac{Cov(\epsilon, D)}{Var(\epsilon)} = \frac{Cov(X\beta, D)}{Var(X\beta)}$$

- When is this assumption reasonable?
  - A number of highly relevant observed control variables are available.
  - Data from high-quality surveys specifically targeted to answer the type of research question you are after.

- The assumption:

$$\frac{\text{Cov}(\epsilon, D)}{\text{Var}(\epsilon)} = \frac{\text{Cov}(X\beta, D)}{\text{Var}(X\beta)}$$

can be used directly to estimate the OLS bias:

$$\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(\epsilon, \tilde{D})}{\text{Var}(\tilde{D})}$$

by substituting:

$$\frac{\text{Cov}(\epsilon, D)}{\text{Var}(\epsilon)} = \frac{\text{Cov}(X\beta, D)}{\text{Var}(X\beta)} \cdot \frac{\text{Var}(\epsilon)}{\text{Var}(\tilde{D})}$$

Estimate the following term step by step:

$$\frac{\text{Cov}(\epsilon, D)}{\text{Var}(\epsilon)} = \frac{\text{Cov}(X\beta, D)}{\text{Var}(X\beta)} \cdot \frac{\text{Var}(\epsilon)}{\text{Var}(\tilde{D})}$$

- Estimate an OLS of outcome on all  $X$  (excluding treatment) – get  $\text{Var}(\epsilon)$  and  $X\hat{\beta}$ .
- Use the predicted index and regress it on treatment,  $D$  – get  $\frac{\text{Cov}(X\beta, D)}{\text{Var}(X\beta)}$ .
- Regress OLS of treatment on all  $X$  – get  $\text{Var}(\tilde{D})$ .

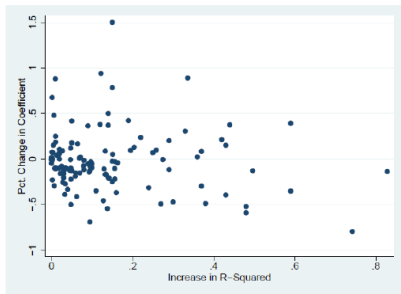


## What Emily Oster adds to AET

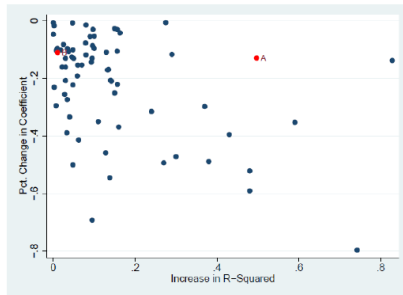
### Critique:

- Coefficient movements alone are not sufficient to calculate bias.
- In particular, this can lead to assuming stability if the included control variable does not explain much of the outcome variance.
- Basic idea is to relate coefficient movements to changes in  $R^2$ .

(a) All Significant Relationships



(b) Sample where Controls Lower Magnitude



Notes: These figures show the relationship between the percent change in coefficient and the increase in R-squared in sample of highly cited papers from top journals in economics. The sample is discussed in Section 2 in Oster (2014).

## Oster's idea:

- Exploit information on coefficients and  $R^2$  values to compute bounds of treatment effect
- Calculate value of proportionality  $\delta$ : captures the explanatory power of unobserved variables as a proportion of the explanatory power of observed variables.
- Quality of the control will be diagnosed by the movement in  $R^2$  when the control is included.

- Calculate an identified set for the treatment effect together with the following bias-adjusted coefficient:

$$\beta' = \tilde{\delta} \frac{(\dot{\beta} - \tilde{\beta})(R_{\max} - \tilde{R})}{(\tilde{R} - \dot{R})}$$

- If  $\tilde{\delta} = 1$

- Calculate model with few basic explanatory variables only!
- Calculate model that controls for a rich set of explanatory variables!
- Look at R-squared - is it considerably higher than in the "raw" model?
- Look at the coefficient of interest - did it stay fairly stable?

- Researchers can either estimate ...  $\delta$ 
  - Assume a value for  $R_{\max}$  and calculate  $\tilde{\delta}$  for which  $\beta = 0$ .
  - Oster (2013) argues that  $\tilde{\delta} = [0, 1]$  is a useful bound.
  - It is possible to assume  $R_{\max} = 1$ .
  - Oster (2013) proposes a rule of thumb to calculate  $R_{\max} = \min\{2.2\tilde{R}, 1\}$ .
- ... or  $\beta$ 
  - Assume a value for  $R_{\max}$  and  $\delta$  to obtain a set of bounds for  $\tilde{\beta}$ .
  - Consider  $\beta^{*'}(R_{\max}, \delta)$ , where  $R_{\max} = \min\{\tilde{R}, 1\}$  and assume that proportional selection is positive:  $\delta = [0, \tilde{\delta}]$ .
  - Define bounds for  $\beta$  where one side of the bound is  $\tilde{\beta}$  and the other side is  $\beta^{*'}(1, \tilde{\delta})$ .

- A lot depends on the hypothetical  $R_{max}$
- Original NBER discussion paper:

$$R_{max} = \min(2.2\tilde{R}, 1)$$

- Updated versions and final publication:

$$R_{max} = \min(1.3\tilde{R}, 1)$$

- Differences in  $R_{max}$  value affects  $\delta$  depending on increase between short and full model.

- A lot depends on the hypothetical  $R_{max}$
- Original NBER discussion paper:

$$R_{max} = \min(2.2\tilde{R}, 1)$$

- Updated versions and final publication:

$$R_{max} = \min(1.3\tilde{R}, 1)$$

- Differences in  $R_{max}$  value affects  $\delta$  depending on increase between short and full model.



- AET and Oster-method try to formalize what many papers do informally: put a number of *stability of estimates*
- This requires remaining assumptions
- Current state of the art: if you need these methods to defend your estimates, you are in a weak situation.
- Referees might ask you to provide Oster-bounds, so you need to be able to do this.