REGRESSION RECAP

Josh Angrist

MIT 14.387 (Fall 2014)

Regression: What You Need to Know

We spend our lives running regressions (I should say: "regressions run me"). And yet this basic empirical tool is often misunderstood. So I begin with a recap of key regression properties. We need these to make sense of IV as well.

Our regression agenda:

1. Three reasons to love
2. The CEF is all you need
3. The long and short of regression anatomy
4. The OVB formula
5. Limited dependent variables and marginal effects
6. Causal vs. casual

The CEF

- The *Conditional Expectation Function* (CEF) for a dependent variable, Y_i given a $\mathrm{K}\times 1$ vector of covariates, X_i (with elements x_{ki}) is written $E[\mathrm{Y}_i|\mathrm{X}_i]$ and is a function of X_i
- Because X_i is random, the CEF is random. For dummy D_i, the CEF takes on two values, $E[\mathrm{Y}_i|\mathrm{D}_i = 1]$ and $E[\mathrm{Y}_i|\mathrm{D}_i = 0]$
- For a specific value of X_i, say $\mathrm{X}_i = 42$, we write $E[\mathrm{Y}_i|\mathrm{X}_i = 42]$
- For continuous Y_i with conditional density $f_y(\cdot|\mathrm{X}_i = x)$, the CEF is

$$E[\mathrm{Y}_i|\mathrm{X}_i = x] = \int t f_y(t|\mathrm{X}_i = x)\, dt$$

  If Y_i is discrete, $E[\mathrm{Y}_i|\mathrm{X}_i = x]$ equals the sum $\sum_t t f_y(t|\mathrm{X}_i = x)$
- The CEF residual is uncorrelated with any function of of X_i. Write $\varepsilon_i \equiv \mathrm{Y}_i - E[\mathrm{Y}_i|\mathrm{X}_i]$.Then for any function, $h(\mathrm{X}_i)$:

$$E[\varepsilon_i h(\mathrm{X}_i)] = E[(\mathrm{Y}_i - E[\mathrm{Y}_i|\mathrm{X}_i])h(\mathrm{X}_i)] = 0$$

  (The LIE proves it)
- **Figure 3.1.1** shows my favorite CEF

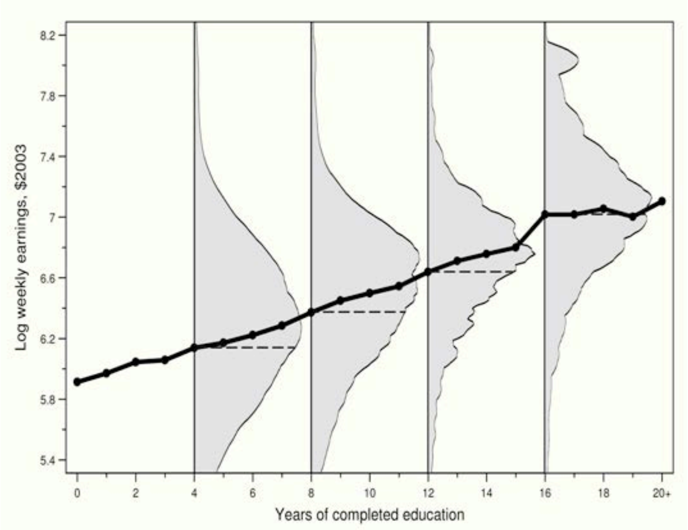


Figure 3.1.1: Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40-49. The data are from the 1980 IPUMS5 percent sample.

Population Regression

- Define *population regression* ("regression," for short) as the solution to the population least squares problem. Specifically, the $\mathrm{K} \times 1$ regression coefficient vector β is defined by solving

$$\beta = \arg\min_b E\left[\left(\mathrm{Y}_i - \mathrm{X}_i' b\right)^2\right]$$

- Using the first-order condition,

$$E\left[\mathrm{X}_i\left(\mathrm{Y}_i - \mathrm{X}_i' b\right)\right] = 0,$$

  the solution for b can be written

$$\beta = E\left[\mathrm{X}_i \mathrm{X}_i'\right]^{-1} E\left[\mathrm{X}_i \mathrm{Y}_i\right]$$

- By construction, $E\left[\mathrm{X}_i\left(\mathrm{Y}_i - \mathrm{X}_i'\beta\right)\right] = 0$. In other words, the population residual, defined as $\mathrm{Y}_i - \mathrm{X}_i'\beta = e_i$, is uncorrelated with the regressors, X_i
- This error term has no life of its own: e_i owes its meaning and existence to β

Three reasons to love

1. Regression solves the population least squares problem and is therefore the BLP of Y_i given X_i
2. If the CEF is linear, regression is it
3. Regression gives the best linear approximation to the CEF

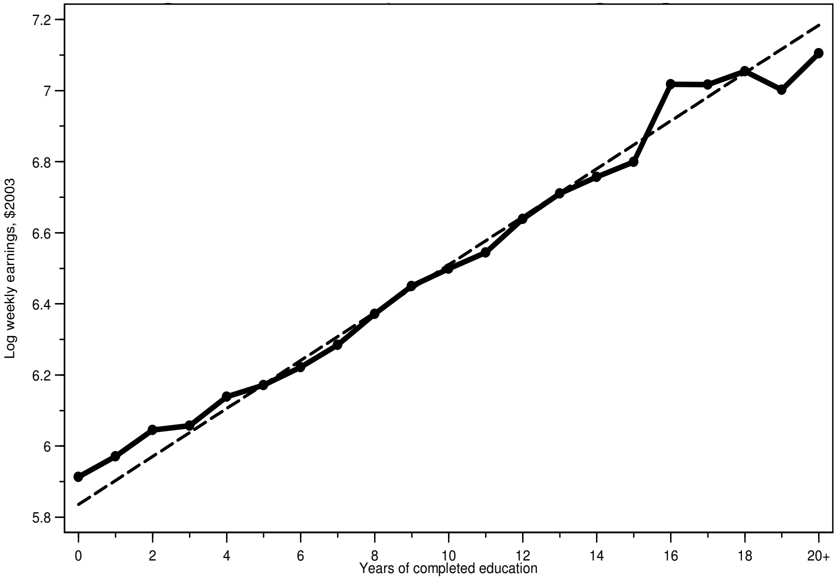
- The first is true by definition; the second follows immediately from CEF-orthgonality. Let's prove the third - it's my favorite!

Theorem

The Regression-CEF Theorem (MHE 3.1.6)
The population regression function $X_i'\beta$ provides the MMSE linear approximation to $E[Y_i|X_i]$, that is,

$$\beta = \arg\min_b E\{(E[Y_i|X_i] - X_i'b)^2\}.$$

- **Figure 3.1.2** illustrates the theorem (What does this depend on?)



Sample is limited to white men, age 40-49. Data is from Census IPUMS 1980, 5% sample.

Figure 3.1.2: Regression threads the CEF of average weekly wages given schooling

The CEF is all you need

- The regression-CEF theorem implies we can use $E[\mathrm{Y}_i|\mathrm{X}_i]$ as a dependent variable instead of Y_i (but watch the weighting!)
- Another way to see this:

$$\beta = E[\mathrm{X}_i\mathrm{X}_i']^{-1}E[\mathrm{X}_i\mathrm{Y}_i] = E[\mathrm{X}_i\mathrm{X}_i']^{-1}E[\mathrm{X}_iE(\mathrm{Y}_i|\mathrm{X}_i)] \quad (1)$$

  The CEF or grouped-data version of the regression formula is useful when working on a project that precludes the analysis of micro data
- To illustrate, we can estimate the schooling coefficient in a wage equation using 21 conditional means, the sample CEF of earnings given schooling
- As **Figure 3.1.3** shows, grouped data weighted by the number of individuals at each schooling level produces coefficients *identical* to that generated by the underlying micro data

A - Individual-level data

```
. regress earnings school, robust

      Source |       SS       df       MS         Number of obs =  409435
-------------+------------------------------      F(  1,409433) =49118.25
       Model | 22631.4793      1  22631.4793      Prob > F      =  0.0000
    Residual |  188648.31 409433  .460755019      R-squared     =  0.1071
-------------+------------------------------      Adj R-squared =  0.1071
       Total | 211279.789 409434   .51602893      Root MSE      =  .67879
-------------+----------------------------------------------------------
             |              Robust                   Old Fashioned
    earnings |      Coef.   Std. Err.      t            Std. Err.       t
-------------+----------------------------------------------------------
      school |    .0674387   .0003447   195.63          .0003043   221.63
      const. |    5.835761   .0045507  1282.39          .0040043  1457.38
------------------------------------------------------------------------
```

B - Means by years of schooling

```
. regress average_earnings school [aweight=count], robust
(sum of wgt is   4.0944e+05)

      Source |       SS       df       MS         Number of obs =      21
-------------+------------------------------      F(  1,    19) =  540.31
       Model |  1.16077332     1  1.16077332      Prob > F      =  0.0000
    Residual |  .040818796    19  .002148358      R-squared     =  0.9660
-------------+------------------------------      Adj R-squared =  0.9642
       Total |  1.20159212    20  .060079606      Root MSE      =  .04635
-------------+----------------------------------------------------------
     average |              Robust                   Old Fashioned
   _earnings |      Coef.   Std. Err.      t            Std. Err.       t
-------------+----------------------------------------------------------
      school |    .0674387   .0040352    16.71         .0029013     23.24
      const. |    5.835761   .0399452   146.09         .0381792    152.85
------------------------------------------------------------------------
```

Figure 3.1.3: Micro-data and grouped-data estimates of returns to schooling. Source: 1980 Census - IPUMS, 5 percent sample. Sample is limited to white men, age 40-49. Derived from Stata regression output. Old-fashioned standard errors are the default reported. Robust standard errors are heteroscedasticity-consistent. Panel A uses individual-level data. Panel B uses earnings averaged by years of schooling.

Regression anatomy lesson

- Bivariate reg recap: the slope coefficient is $\beta_1 = \frac{Cov(\mathrm{Y}_i, x_i)}{V(x_i)}$, and the intercept is $\alpha = E[\mathrm{Y}_i] - \beta_1 E[\mathrm{X}_i]$
- With more than one non-constant regressor, the k-th non-constant slope coefficient is:

$$\beta_k = \frac{Cov(\mathrm{Y}_i, \tilde{x}_{ki})}{V(\tilde{x}_{ki})}, \tag{2}$$

  where $\tilde{x}_{ki}$ is the residual from a regression of x_{ki} on all other covariates
- The anatomy formula shows us that each coefficient in a multivariate regression is the bivariate slope coefficient for the corresponding regressor, after "partialing out" other variables in the model.
- Verify the regression-anatomy formula by subbing

$$\mathrm{Y}_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \ldots + \beta_{\mathrm{K}} x_{\mathrm{K}i} + e_i$$

  in the numerator of (2) and work through to find that $Cov(\mathrm{Y}_i, \tilde{x}_{ki}) = \beta_k V(\tilde{x}_{ki})$

Omitted Variables Bias

- The omitted variables bias (OVB) formula describes the relationship between regression estimates in models with different controls
- Go long: wages on schooling, S_i, controlling for ability (A_i)

$$\mathrm{Y}_i = \alpha + \rho \mathrm{S}_i + A_i' \gamma + \varepsilon_i \tag{3}$$

- Ability is hard to measure. What if we leave it out? The result is

$$\frac{Cov(\mathrm{Y}_i, \mathrm{S}_i)}{V(\mathrm{S}_i)} = \rho + \gamma' \delta_{As},$$

where δ_{As} is the vector of coefficients from regressions of the elements of A_i on S_i . . .

  - *Short equals long plus the effect of omitted times the regression of omitted on included*

- Short equals long when omitted and included are uncorrelated
- **Table 3.2.1** illustrates OVB (some controls are bad; the formula works for good and bad alike)

TABLE 3.2.1
Estimates of the returns to education for men in the NLSY

Controls:	(1) None	(2) Age Dummies	(3) Col. (2) and Additional Controls*	(4) Col. (3) and AFQT Score	(5) Col. (4), with Occupation Dummies
	.132	.131	.114	.087	.066
	(.007)	(.007)	(.007)	(.009)	(.010)

Notes: Data are from the National Longitudinal Survey of Youth (1979 cohort, 2002 survey). The table reports the coefficient on years of schooling in a regression of log wages on years of schooling and the indicated controls. Standard errors are shown in parentheses. The sample is restricted to men and weighted by NLSY sampling weights. The sample size is 2,434.

*Additional controls are mother's and father's years of schooling, and dummy variables for race and census region.

Limited dependent variables

- Regression always make sense in the sense that regression approximates the CEF
- Can I *really* use OLS if my dependent variable is . . . a dummy (like employment); non-negative (like earnings); a count variable (like weeks worked)?
- *Regress easy, grasshopper* . . . but if you do stray, show me the MFX
- Probing probit: assume that LFP is determined by a latent variable, Y_i^*, satisfying

$$\mathrm{Y}_i^* = \beta_0^* + \beta_1^* \mathrm{S}_i - \nu_i, \tag{4}$$

where ν_i is distributed $N(0, \sigma_\nu^2)$. The latent index model says

$$\mathrm{Y}_i = 1[\mathrm{Y}_i^* > 0],$$

so the CEF can be written

$$E[\mathrm{Y}_i|\mathrm{S}_i] = \Phi\left[\frac{\beta_0^* + \beta_1^* \mathrm{S}_i}{\sigma_\nu}\right]$$

Limited dependent variables (cont.)

- For Bernoulli S_i:

$$E[\mathrm{Y}_i|\mathrm{S}_i] = \Phi\left[\frac{\beta_0^*}{\sigma_\nu}\right] + \left\{\Phi\left[\frac{\beta_0^*+\beta_1^*}{\sigma_\nu}\right] - \Phi\left[\frac{\beta_0^*}{\sigma_\nu}\right]\right\}\mathrm{S}_i$$

  OLS is bang on here! (why?)

- But it ain't always about treatment effects; MFX for probit are

$$E\left(\frac{\partial E[\mathrm{Y}_i|\mathrm{S}_i]}{\partial \mathrm{S}_i}\right) = E\left(\varphi\left[\frac{\beta_0^*+\beta_1^*\mathrm{S}_i}{\sigma_\nu}\right]\right)\frac{\beta_1^*}{\sigma_\nu} \tag{5}$$

  Index coefficients tell us only the sign of the effect of S_i on average Y_i (sometimes, as in MNL, not even that)

- For logit:

$$E\left(\frac{\partial E[\mathrm{Y}_i|\mathrm{S}_i]}{\partial \mathrm{S}_i}\right) = E[\Lambda(\beta_0^*+\beta_1^*\mathrm{S}_i)(1-(\Lambda\beta_0^*+\beta_1^*\mathrm{S}_i))]\beta_1^* \tag{6}$$

- OLS and MFX from any nonlinear alternative are usually close (identical for probit when S_i is Normal)

Making MFX

- Are derivative-based MFX kosher in a discrete-regressor scenario?
  - With covariates, stata generates discrete average derivatives like this,

$$E\left\{\Phi\left[\frac{\beta_0^{*\prime}X_i+\beta_1^*}{\sigma_\nu}\right]-\Phi\left[\frac{\beta_0^{*\prime}X_i}{\sigma_\nu}\right]\right\} \tag{7}$$

  - Note that

$$\Phi\left[\frac{X_i'\beta_0^*+\beta_1^*}{\sigma_\nu}\right]=\Phi\left[\frac{X_i'\beta_0^*}{\sigma_\nu}\right]+\varphi\left[\frac{X_i'\beta_0^*+\Delta_i}{\sigma_\nu}\right]\beta_1^*$$

    for some $\Delta_i \in [0, \beta_1^*]$. So the continuous MFX calculation

$$E\left\{\varphi\left[\frac{X_i'\beta_0^*+\beta_1^*\mathrm{s}_i}{\sigma_\nu}\right]\right\}\beta_1^* \tag{8}$$

    approximates the discrete

- Stata notices discrete regressors, in which case you' llget (7) unless you ask for (8)
- OLS vindicated: MHE **Table 3.4.2**

Table 3.4.2

Comparison of alternative estimates of the effect of childbearing on LDVs

		Right-Hand-Side Variable								
		More than Two Children					Number of Children			
			Probit		Tobit			Probit MFX	Tobit MFX	
Dependent variable	Mean (1)	OLS (2)	Avg. Effect, Full Sample (3)	Avg. Effect on Treated (4)	Avg. Effect, Full Sample (5)	Avg. Effect on Treated (6)	OLS (7)	Avg. Effect, Full Sample (8)	Avg. Effect, Full Sample (9)	Avg. Effect on Treated (10)
A. Full sample										
Employment	.528	−.162	−.163	−.162	—	—	−.113	−.114	—	—
	(.499)	(.002)	(.002)	(.002)			(.001)	(.001)		
Hours worked	16.7	−5.92	—	—	−6.56	−5.87	−4.07	—	−4.66	−4.23
	(18.3)	(.074)			(.081)	(.073)	(.047)		(.054)	(.049)
B. Nonwhite college attenders over age 30, first birth before age 20										
Employment	.832	−.061	−.064	−.070	—	—	−.054	−.048	—	—
	(.374)	(.028)	(.028)	(.031)			(.016)	(.013)		
Hours worked	30.8	−4.69	—	—	−4.97	−4.90	−2.83	—	−3.20	−3.15
	(16.0)	(1.18)			(1.33)	(1.31)	(.645)		(.670)	(.659)

Notes: The table reports OLS estimates, average treatment effects, and marginal effects (MFX) for the effect of childbearing on mothers' labor supply. The sample in panel A includes 254,654 observations and is the same as the 1980 census sample of married women used by Angrist and Evans (1998). Covariates include age, age at first birth, and dummies for boys at first and second birth. The sample in panel B includes 746 nonwhite women with at least some college aged over 30 whose first birth was before age 20. Standard deviations are reported in parentheses in column 1. Standard errors are shown in parentheses in other columns. The sample used to estimate average effects on the treated in columns 4, 6, and 10 includes women with more than two children.

Casual vs. causal

- *Casual* regressions happen for many reasons: exploratory or descriptive analysis, just having fun, no long-term commitment . . .
- *Causal* regressions are more serious and enduring, describe counterfactual states of the world, useful for policy analysis
- Americans mortgage homes to send a child to elite private colleges. Does private pay? Denote private attendance by C_i. The causal relationship between private college attendance and earnings is

$$\begin{array}{ll} \mathrm{Y}_{1i} & \text{if } \mathrm{C}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{C}_i = 0 \end{array}$$

- $\mathrm{Y}_{1i} - \mathrm{Y}_{0i}$ is an individual causal effect. Alas, we only get to see one of Y_{1i} or Y_{0i}. The observed outcome, Y_i, is

$$\mathrm{Y}_i = \mathrm{Y}_{0i} + (\mathrm{Y}_{1i} - \mathrm{Y}_{0i})\mathrm{C}_i \tag{9}$$

We hope to measure average $\mathrm{Y}_{1i} - \mathrm{Y}_{0i}$ for some group, say those who went private: $E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{C}_i = 1]$, i.e., TOT

Casual vs. causal (cont.)

- Comparisons of those who did and didn't go private are biased:

$$\underbrace{E\left[\mathrm{Y}_i|\mathrm{C}_i=1\right]-E[\mathrm{Y}_i|\mathrm{C}_i=0]}_{\text{Observed difference in earnings}} = \underbrace{E[\mathrm{Y}_{1i}-\mathrm{Y}_{0i}|\mathrm{C}_i=1]}_{\text{TOT}} \tag{10}$$

$$+\underbrace{E\left[\mathrm{Y}_{0i}|\mathrm{C}_i=1\right]-E\left[\mathrm{Y}_{0i}|\mathrm{C}_i=0\right]}_{\text{selection bias}}$$

- It seems likely that those who go to private college would have earned more anyway. The naive comparison, $E\left[\mathrm{Y}_i|\mathrm{C}_i=1\right]-E[\mathrm{Y}_i|\mathrm{C}_i=0]$, exaggerates the benefits of private college attendance
    - *Selection bias = OVB in a causal model*
- The *conditional independence assumption* (CIA) asserts that conditional on observed X_i, selection bias disappears:

$$\{\mathrm{Y}_{0i},\mathrm{Y}_{1i}\} \amalg \mathrm{C}_i|\mathrm{X}_i \tag{11}$$

- Given the CIA, conditional-on-X_i comparisons are causal:

$$E\left[\mathrm{Y}_i|\mathrm{X}_i,\mathrm{C}_i=1\right]-E\left[\mathrm{Y}_i|\mathrm{X}_i,\mathrm{C}_i=0\right]=E[\mathrm{Y}_{1i}-\mathrm{Y}_{0i}|\mathrm{X}_i]$$

Using the CIA

- The CIA means that C_i is "as good as randomly assigned," conditional on X_i
- A secondary implication: Given the CIA, the conditional on X_i causal effect of private college attendance on private graduates equals the average private effect at X_i:

$$E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{X}_i, \mathrm{C}_i = 1\right] = E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{X}_i\right]$$

- This is important . . . but less important than the elimination of selection bias
- Note also that the marginal average private college effect can be obtained by averaging over X_i:

$$\begin{aligned} & E\{E\left[\mathrm{Y}_i | \mathrm{X}_i, \mathrm{C}_i = 1\right] - E\left[\mathrm{Y}_i | \mathrm{X}_i, \mathrm{C}_i = 0\right]\} \\ = \;& E\{E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | \mathrm{X}_i]\} \\ = \;& E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}] \end{aligned}$$

- This suggests we compare people with the same X's ... like matching . . . but I wanna regress!

Regression and the CIA

- The regression machine turns the CIA into causal effects
- Constant causal effects allow us to focus on selection issues (MHE 3.3 relaxes this). Suppose

$$\begin{aligned} Y_{0i} &= \alpha + \eta_i \\ Y_{1i} &= Y_{0i} + \rho \end{aligned} \tag{12}$$

- Using (9) and (12), we have

$$\mathrm{Y}_i = \alpha + \rho \mathrm{C}_i + \eta_i \tag{13}$$

- Equation (13) *looks* like a bivariate regression model, except that (12) associates the coefficients in (13) with a causal relationship
- This is not a regression, because C_i can be correlated with potential outcomes, in this case, the residual, η_i

Regression and the CIA (cont.)

- The CIA applied to our constant-effects setup implies:

$$E[\eta_i|\mathrm{C}_i, \mathrm{X}_i] = E[\eta_i|\mathrm{X}_i]$$

- Suppose also that

$$E[\eta_i|\mathrm{X}_i] = \mathrm{X}_i'\gamma$$

  so that

$$E[\mathrm{Y}_i|\mathrm{X}_i, \mathrm{C}_i] = \alpha + \rho\mathrm{C}_i + E[\eta_i|\mathrm{X}] = \alpha + \rho\mathrm{C}_i + \mathrm{X}_i'\gamma$$

- Mean-independence implies orthogonality, so

$$\mathrm{Y}_i = \alpha + \rho\mathrm{C}_i + \mathrm{X}_i'\gamma + v_i \tag{14}$$

  has error

$$v_i \equiv \eta_i - \mathrm{X}_i'\gamma = \eta_i - E[\eta_i|\mathrm{C}_i, \mathrm{X}_i]$$

  uncorrelated with regressors, C_i and X_i. The same ρ appears in the regression and causal models!

- Modified **Dale and Krueger (2002)**: private proving ground

Matchmaker, Matchmaker . . . Find Me a College!

		Private			Public			
Applicant Group	Student	Ivy	Leafy	Smart	All State	Ball State	Altered State	1996 Earnings
A	1		Reject	**Admit**		Admit		110,000
	2		Reject	**Admit**		Admit		100,000
	3		Reject	Admit		**Admit**		110,000
B	4	**Admit**			Admit		Admit	60,000
	5	Admit			Admit		**Admit**	30,000
C	6		**Admit**					115,000
	7		**Admit**					75,000
D	8	Reject			**Admit**	Admit		90,000
	9	Reject			Admit	**Admit**		60,000

Notes: Students enroll at the college indicated in **bold**; enrollment decisions are also highlighted in grey.

Table 2.1: The College Matching Matrix

	No Selection Controls			Selection Controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private School	0.135	0.095	0.086	0.007	0.003	0.013
	(0.055)	(0.052)	(0.034)	(0.038)	(0.039)	(0.025)
Own SAT score/100		0.048	0.016		0.033	0.001
		(0.009)	(0.007)		(0.007)	(0.007)
Predicted log(Parental Income)			0.219			0.190
			(0.022)			(0.023)
Female			-0.403			-0.395
			(0.018)			(0.021)
Black			0.005			-0.040
			(0.041)			(0.042)
Hispanic			0.062			0.032
			(0.072)			(0.070)
Asian			0.170			0.145
			(0.074)			(0.068)
Other/Missing Race			-0.074			-0.079
			(0.157)			(0.156)
High School Top 10 Percent			0.095			0.082
			(0.027)			(0.028)
High School Rank Missing			0.019			0.015
			(0.033)			(0.037)
Athlete			0.123			0.115
			(0.025)			(0.027)
Selection Controls	N	N	N	Y	Y	Y

Notes: Columns (1)-(3) include no selection controls. Columns (4)-(6) include a dummy for each group formed by matching students according to schools at which they were accepted or rejected. Each model is estimated using only observations with Barron's matches for which different students attended both private and public schools. The sample size is 5,583. Standard errors are shown in parentheses.

Table 2.2: Private School Effects: Barron's Matches

	No Selection Controls			Selection Controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private School	0.212	0.152	0.139	0.034	0.031	0.037
	(0.060)	(0.057)	(0.043)	(0.062)	(0.062)	(0.039)
Own SAT Score/100		0.051	0.024		0.036	0.009
		(0.008)	(0.006)		(0.006)	(0.006)
Predicted log(Parental Income)			0.181			0.159
			(0.026)			(0.025)
Female			-0.398			-0.396
			(0.012)			(0.014)
Black			-0.003			-0.037
			(0.031)			(0.035)
Hispanic			0.027			0.001
			(0.052)			(0.054)
Asian			0.189			0.155
			(0.035)			(0.037)
Other/Missing Race			-0.166			-0.189
			(0.118)			(0.117)
High School Top 10 Percent			0.067			0.064
			(0.020)			(0.020)
High School Rank Missing			0.003			-0.008
			(0.025)			(0.023)
Athlete			0.107			0.092
			(0.027)			(0.024)
Average SAT Score of				0.110	0.082	0.077
Schools Applied to/100				(0.024)	(0.022)	(0.012)
Sent Two Application				0.071	0.062	0.058
				(0.013)	(0.011)	(0.010)
Sent Three Applications				0.093	0.079	0.066
				(0.021)	(0.019)	(0.017)
Sent Four or more Applications				0.139	0.127	0.098
				(0.024)	(0.023)	(0.020)

Note: Standard errors are shown in parentheses. The sample size is 14,238.

Table 2.3: Private School Effects: Average SAT Controls

	No Selection Controls			Selection Controls		
	(1)	(2)	(3)	(4)	(5)	(6)
School Avg. SAT Score/100	0.109	0.071	0.076	-0.021	-0.031	0.000
	(0.026)	(0.025)	(0.016)	(0.026)	(0.026)	(0.018)
Own SAT score/100		0.049	0.018		0.037	0.009
		(0.007)	(0.006)		(0.006)	(0.006)
Predicted log(Parental Income)			0.187			0.161
			(0.024)			(0.025)
Female			-0.403			-0.396
			(0.015)			(0.014)
Black			-0.023			-0.034
			(0.035)			(0.035)
Hispanic			0.015			0.006
			(0.052)			(0.053)
Asian			0.173			0.155
			(0.036)			(0.037)
Other/Missing Race			-0.188			-0.193
			(0.119)			(0.116)
High School Top 10 Percent			0.061			0.063
			(0.018)			(0.019)
High School Rank Missing			0.001			-0.009
			(0.024)			(0.022)
Athlete			0.102			0.094
			(0.025)			(0.024)
Average SAT Score of				0.138	0.116	0.089
Schools Applied To/100				(0.017)	(0.015)	(0.013)
Sent Two Application				0.082	0.075	0.063
				(0.015)	(0.014)	(0.011)
Sent Three Applications				0.107	0.096	0.074
				(0.026)	(0.024)	(0.022)
Sent Four or more Applications				0.153	0.143	0.106
				(0.031)	(0.030)	(0.025)

Note: Standard errors are shown in parentheses. The sample size is 14,238.

Table 2.4: School Selectivity Effects: Average SAT Controls

	Dependent Variable					
	Own SAT score/100			Predicted log(Parental Income)		
	(1)	(2)	(3)	(4)	(5)	(6)
Private School	1.165	1.130	0.066	0.128	0.138	0.028
	(0.196)	(0.188)	(0.112)	(0.035)	(0.037)	(0.037)
Female		-0.367			0.016	
		(0.076)			(0.013)	
Black		-1.947			-0.359	
		(0.079)			(0.019)	
Hispanic		-1.185			-0.259	
		(0.168)			(0.050)	
Asian		-0.014			-0.060	
		(0.116)			(0.031)	
Other/Missing Race		-0.521			-0.082	
		(0.293)			(0.061)	
High School Top 10 Percent		0.948			-0.066	
		(0.107)			(0.011)	
High School Rank Missing		0.556			-0.030	
		(0.102)			(0.023)	
Athlete		-0.318			0.037	
		(0.147)			(0.016)	
Average SAT Score of			0.777			0.063
Schools Applied To/100			(0.058)			(0.014)
Sent Two Application			0.252			0.020
			(0.077)			(0.010)
Sent Three Applications			0.375			0.042
			(0.106)			(0.013)
Sent Four or more Applications			0.330			0.079
			(0.093)			(0.014)

Note: Standard errors are shown in parentheses. The sample size is 14,238.

Table 2.5: Private School Effects: Omitted Variable Bias

What Next?

- Regression always makes sense ... in the sense that it provides best-in-class approximation to the CEF
- MFX from more elaborate non-linear models are usually indistinguishable from the corresponding regression estimates
- We're not always content to run regressions, of course, though this is usually where we start
  - Regression is our first line of attack on the identification problem; it's all about *control*
- If the regression you've got is not the one you want, that's because the underlying *relationship* is unsatisfactory
- Whats to be done with an unsatisfactory relationship?
  - Move on, grasshopper ... to IV!
- But wait: we need some *training* first