1. Regression and causality

LPO 8852: Regression II

Sean P. Corcoran

What you learned in Regression I

The mechanics and properties of linear regression models:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + u_i$$

- Model specification
- Estimation (e.g., OLS, WLS)
- Inference: What is the *standard error* of your estimator? What is the estimator's *sampling distribution* in finite samples? In large samples? Knowledge of the sampling distribution is needed to construct *confidence intervals* and conduct *hypothesis tests*.

Model interpretation and statistical inference rely heavily on assumptions.

What you learned in Regression I

When I first learned econometrics, I often felt dissatisfied:

- Assumptions feel implausible
- How do we know the model is "correct"?
- There are always "omitted variables"!
- Causal interpretation feels like a pipe dream.

Regression II

Research designs for causal inference

- When can a regression be interpreted as causal?
- What does it mean for an estimator to have a causal interpretation?
- What research designs—which may or may not use regression—make a strong case for causal interpretation?

We will consider:

- Matching estimators
- Difference-in-differences
- Other panel data models
- Instrumental variables
- Regression discontinuity
- Synthetic control methods

What is regression and what is it good for?

We often use regression to estimate the *conditional expectation function* (CEF) for Y_i given values of one or more other variables $X_{i1}, X_{i2}, ..., X_{iK}$. That is, we seek to estimate parameters of a function that tell us the <u>mean</u> of Y given specific values of X ($E[Y_i|X_{i1}, X_{i2}, ..., X_{iK}]$). Importantly:

- The CEF need not be *linear*
- The CEF need not be *causal*

The CEF is a population concept. We typically use sample data to estimate it.

Conditional expectation functions

From Angrist & Pischke (2009): CEF of log weekly wages given years of completed schooling

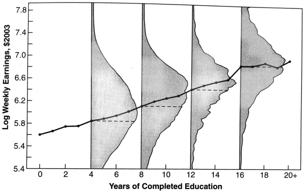

Figure 3.1.1 Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40–49 in the 1980 IPUMS 5 percent file.

Conditional expectation functions

For simplicity, consider only one predictor variable X_i. We can decompose Y_i into two parts, the CEF and an error term: $Y_i = E[Y_i|X_i] + \epsilon_i$, where:

- ϵ_i is mean independent of X, that is $E[\epsilon_i|X_i] = 0$ and
- ϵ_i is uncorrelated with any function of X

In other words, the CEF fully captures the relationship between Y and X.

The decomposition of Y_i is into a piece "explained by X_i" and a leftover orthogonal (uncorrelated) piece.

Conditional expectation functions

Why do we care about conditional expectation functions?

- They are a good summary of the relationship between Y and X. Why? We think of means as representative values.
- The relationship between Y and X may be useful for *prediction*, in a statistical sense.
- The CEF is the *best predictor* of Y given X in that it minimizes the sum of squared errors (ϵ) in the population.
- They sometimes describe a causal relationship.

Even a CEF that is *not* causal can be useful.

Linear regression

The population CEF is an unknown function. In practice, we typically estimate a *population regression function*—the *line* that best fits the population distribution of (Y_i, X_i).

- Simple: $Y_i = \beta_0 + \beta_1 X_i + u_i$
- Multiple: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_3 X_{ki} + u_i$

In the simple regression case, the least squares slope and intercept—those that minimize the sum of squared errors in the population—are:

$$\beta_1 = \frac{Cov(Y_i, X_i)}{Var(X_i)}$$

$$\beta_0 = E[Y_i] - \beta_1 E[X_i]$$

Linear regression

What good is the population regression function?

1. If the CEF happens to be linear, then the PRF is the CEF. This is unlikely in most real world-cases but true in two special cases: joint normality, and *saturated* regression models.
2. The PRF is the best *linear* predictor of Y_i given the X_i.
3. The PRF provides the least squares approximation to the CEF when the CEF is nonlinear.

#3 above is key. Even when the CEF is nonlinear, the PRF is the best linear approximation to it.

Linear regression

From Angrist & Pischke (2009): linear regression as an approximation to the CEF

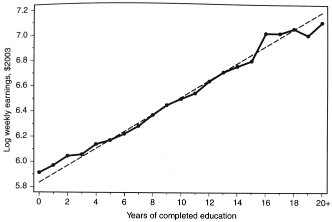

Figure 3.1.2 Regression threads the CEF of average weekly wages given schooling (dots = CEF; dashes = regression line).

Saturated regression models

A *saturated* regression model is a model with discrete explanatory variables that includes a separate parameter for every possible combination of values taken by the explanatory variables.

Saturated regression models

Example: two dummy (0/1) explanatory variables X_1 and X_2

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma_1 X_{1i} X_{2i} + u_i$$

There are four possible combinations of X_1 and X_2 and thus four possible predictions of $Y|X$:

X_1	X_2	$E(Y\|X)$
0	0	β_0
1	0	$\beta_0 + \beta_1$
0	1	$\beta_0 + \beta_2$
1	1	$\beta_0 + \beta_1 + \beta_2 + \gamma_1$

The coefficients are *main effects* (β_1, β_2) and an *interaction term* (γ_1).

Saturated regression models

Estimating the above PRF is equivalent to estimating the CEF $E[Y|X_1, X_2]$. The PRF fits the CEF perfectly—since there is no other way the CEF can be specified.

Regression and causality

The PRF is useful as a "best approximation" to the population CEF. But its slope coefficients are *not necessarily causal*. So when will regression have a causal interpretation?

A regression is causal when the CEF it approximates is causal. (Angrist & Pischke, 2009).

A CEF is causal when it describes differences in *average potential outcomes* for a given reference population. What does this mean?

Causality

A *causal effect* is a change in some feature of the world that would result from a change to some other feature of the world.

It involves a *counterfactual comparison* between the actual world and a hypothetical world in which there was no change in the feature of the world claimed to have a causal effect.

Potential outcomes are useful for thinking about counterfactuals. The potential outcomes framework was introduced by Neyman (1923) and later generalized by Rubin in the 1970s and 1980s. This framework is often called the *Rubin causal model*.

Potential outcomes

Let D_i be a dichotomous indicator of a "treatment" where $D_i = 1$ means unit i is "treated" and $D_i = 0$ means i is "not treated." For every i there are two potential outcomes:

- $Y_i(1)$ or Y_{i1} = outcome when $D = 1$
- $Y_i(0)$ or Y_{i0} = outcome when $D = 0$

These are referred to as *potential outcomes* since units are not observed in more than one state.

Potential outcomes

This is the "fundamental problem of causal inference." The observed Y_i is either $Y_i(0)$ or $Y_i(1)$:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

A *counterfactual* is the outcome for the unit in the other (hypothetical, unobserved) state. E.g., the counterfactual for treated i would be $Y_i(0)$.

Example 1: job training program

Person	D_i	$Y_i(0)$	$Y_i(1)$	Y_i
1	1	10	14	14
2	1	8	12	12
3	1	12	16	16
4	1	8	12	12
5	1	6	10	10
6	1	4	8	8
7	0	4	8	4
8	0	6	10	6
9	0	8	12	8
10	0	4	8	4
11	0	10	14	10
12	0	8	12	8
13	0	2	6	2
14	0	1	5	1
Mean	0.429	6.5	10.5	8.2

Source: Jennifer Hill (2011) lecture notes. Assume Y_i is earnings and D_i indicates participation in job training program.

Treatment effects

The causal effect of D on Y for individual i (the *treatment effect*) is:

$$\tau_i = Y_i(1) - Y_i(0)$$

We are often interested in the population average treatment effect (ATE):

$$ATE = E(\tau) = E[\underbrace{Y(1) - Y(0)}_{\text{not observed}}]$$

Or the average treatment effect on the treated (ATT):

$$ATT = E(\tau|D=1) = E[Y(1)|D=1] - \underbrace{E[Y(0)|D=1]}_{\text{not observed}}$$

Treatment effects

Or the average treatment effect on the untreated (ATU):

$$ATU = E(\tau|D=0) = \underbrace{E[Y(1)|D=0]}_{\text{not observed}} - E[Y(0)|D=0]$$

The ATE, ATT, and ATU are *estimands*—quantities of interest in the population. Researchers are often most interested in ATT or ATE.

Note the ATE is a weighted average of the ATT and ATU:

$$ATE = pATT + (1-p)ATU$$

where p is the proportion treated.

Treatment effects

Suppose we compare means for the treated and untreated:

$$\begin{aligned} E[Y(1)|D=1] - E[Y(0)|D=0] = {} & E[Y(1)|D=1] - E[Y(0)|D=0] \\ & - \underbrace{E[Y(0)|D=1] + E[Y(0)|D=1]}_{0} \end{aligned}$$

$$E[Y(1)|D=1] - E[Y(0)|D=0] = ATT + \underbrace{E[Y(0)|D=1] - E[Y(0)|D=0]}_{\text{selection bias}}$$

Selection bias reflects "baseline differences" in $Y(0)$ between the treated and untreated group.

Example 1: job training program

Person	D_i	Educ.	Age	$Y(0)$	$Y(1)$	Y
1	1	1	26	10	14	14
2	1	1	21	8	12	12
3	1	1	30	12	16	16
4	1	1	19	8	12	12
5	1	0	25	6	10	10
6	1	0	22	4	8	8
Mean ($D=1$)	1	0.67	23.8	8	12	12
7	0	0	21	4	8	4
8	0	0	26	6	10	6
9	0	0	28	8	12	8
10	0	0	20	4	8	4
11	0	1	26	10	14	10
12	0	1	21	8	12	8
13	0	0	16	2	6	2
14	0	0	15	1	5	1
Mean ($D=0$)	0	0.25	21.6	5.4	9.4	5.4

Treatment effects

In Example 1, $ATT = 4$. But:

$$E[Y(1)|D=1]-E[Y(0)|D=0] = ATT + \underbrace{E[Y(0)|D=1] - E[Y(0)|D=0]}_{\text{selection bias}}$$

$$12.0 - 5.4 = 4.0 + \underbrace{8.0 - 5.4}_{\text{selection bias}}$$

The treated group has a higher $Y(0)$ than the untreated group. This could be due to their higher average education and age, two things associated with higher earnings. Their Y would have been higher on average even in the absence of treatment.

Treatment effects

Note the difference in group means also fails to recover the ATE:

$$E[Y(1)|D=1] - E[Y(0)|D=0] = ATE + \underbrace{E[Y(0)|D=1] - E[Y(0)|D=0]}_{\text{selection bias}} + \underbrace{(1-p)(ATT - ATU)}_{\text{heterogeneous treatment effect bias}}$$

See *Mixtape* chapter on potential outcomes for the algebra.

Heterogeneous treatment effects

In Example 1, $ATT = ATE$. In practice, ATT often differs from the ATE because units endogeneously sort into treatments based on gains they expect from it.

Regression and causality

What does this have to do with regression? We often use regression to estimate the ATE or ATT. Suppose we estimate the following population regression function for Example 1:

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

- Note the PRF is the CEF in this case—it's a saturated model and gives us the mean of Y_i for given values of D_i (0 or 1).
- Can this CEF be interpreted as causal?
- Does it describe differences in *average potential outcomes* for a given reference population?

Regression and causality

In Example 1 there are constant treatment effects. For every i, $Y_i(1) = Y_i(0) + \delta$. In the population:

$$\begin{aligned}\beta_1 &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &= E[Y_i(0)|D_i = 1] + \delta - E[Y_i(0)|D_i = 0] \\ &= \delta + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}\end{aligned}$$

The CEF here does not describe differences in average potential outcomes for any reference population. β_1 doesn't correspond to either ATE or ATT.

The experimental ideal

Under what conditions will selection bias be zero? When treatment assignment is *independent* of potential outcomes:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D$$

One case where this holds is *randomization* to treatment. Under random assignment, $E[Y_i(0)|D_i = 1] = E[Y_i(0)|D_i = 0]$. In other words, the expected outcome is the same in the absence of treatment.

The two groups are drawn from the same population, so the CEF now describes differences in average potential outcomes for this population.

Conditional independence assumption

Suppose that potential outcomes depend on X_i:

$$Y_i(0) = \alpha_0 + \alpha_1 X_i$$
$$Y_i(1) = \alpha_0 + \alpha_1 X_i + \delta$$

and that there is selection into treatment, such that D_i and X_i are correlated:

$$X_i = \gamma_0 + \gamma_1 D_i$$

We estimate the naive regression:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Conditional independence assumption

As before:

$$\beta_1 = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

$$= \alpha_0 + \alpha_1 E[X_i|D_i = 1] + \delta - \alpha_0 - \alpha_1 E[X_i|D_i = 0]$$

$$= \delta + \underbrace{\alpha_1(E[X_i|D_i = 1] - E[X_i|D_i = 0])}_{\text{selection bias}}$$

$$= \delta + \underbrace{\alpha_1(\gamma_0 + \gamma_1 - \gamma_0)}_{\text{selection bias}}$$

$$= \delta + \underbrace{\alpha_1\gamma_1}_{\text{selection bias}}$$

Conditional independence assumption

Estimating a regression that *conditions on* (controls for) X would eliminate the selection bias. Here, the only reason treated and untreated units differ in their potential outcomes is that they have different levels of X.

Example 2: private colleges

Does attending a selective private college result in higher earnings?

TABLE 2.3
Private school effects: Average SAT score controls

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212 (.060)	.152 (.057)	.139 (.043)	.034 (.062)	.031 (.062)	.037 (.039)
Own SAT score ÷ 100		.051 (.008)	.024 (.006)		.036 (.006)	.009 (.006)
Log parental income			.181 (.026)			.159 (.025)
Female			−.398 (.012)			−.396 (.014)
Black			−.003 (.031)			−.037 (.035)
Hispanic			.027 (.052)			.001 (.054)
Asian			.189 (.035)			.155 (.037)
Other/missing race			−.166 (.118)			−.189 (.117)
High school top 10%			.067 (.020)			.064 (.020)
High school rank missing			.003 (.025)			−.008 (.023)
Athlete			.107 (.027)			.092 (.024)
Average SAT score of schools applied to ÷ 100				.110 (.024)	.082 (.022)	.077 (.012)
Sent two applications				.071 (.013)	.062 (.011)	.058 (.010)
Sent three applications				.093 (.021)	.079 (.019)	.066 (.017)
Sent four or more applications				.139 (.024)	.127 (.023)	.098 (.020)

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column shows coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Source: *Mastering Metrics ch. 2*

Example 2: private colleges

Attendance at a private college is not randomly assigned; we should be concerned that the CEF does not describe differences in average *potential* outcomes. It may be that students attending selective private colleges are better qualified on a number of dimensions than students not attending such colleges.

If the CEF we are estimating does not describe differences in average potential outcomes, we say the causal effect is *not identified*.

Another example: class size

Omitted variables bias

Suppose instead that potential outcomes are described by the following "long" regression, where Y_i is (log) earnings, P_i is an indicator variable for private college attendance and A_i is a measure of "ability":

$$Y_i = \alpha^\ell + \beta^\ell P_i + \gamma A_i + e_i^\ell$$

The "short" regression estimated in column (1) above is:

$$Y_i = \alpha^s + \beta^s P_i + e_i^s$$

We can estimate the "short" regression, but if the true model of potential outcomes is the "long" regression ($\gamma \neq 0$), we may have *omitted variables bias*. The error term in the "short" regression is: $e_i^s = \gamma A_i + e_i^\ell$.

Omitted variables bias

There is a formal (and mechanical) link between β^s and β^ℓ:

$$\beta^s = \beta^\ell + \pi_1 \gamma$$

Where:

- γ comes from the long regression: it is the relationship between A_i and Y_i (conditional on P_i).
- π_1 comes from an "auxiliary" regression of the omitted variable (A_i) on the included variable (P_i).

$$A_i = \pi_0 + \pi_1 P_i + v_i$$

Example

Auxiliary regressions where A_i is the student's SAT score (in hundreds):

TABLE 2.3
Private school effects: Average SAT score controls

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212	.152	.139	.034	.031	.037
	(.060)	(.057)	(.043)	(.062)	(.062)	(.039)
Own SAT score ÷ 100		.051	.024		.036	.009
		(.008)	(.006)		(.006)	(.006)
Log parental income			.181			.159
			(.026)			(.025)
Female			−.398			−.396
			(.012)			(.014)
Black			−.003			−.037
			(.031)			(.035)
Hispanic			.027			.001
			(.052)			(.054)
Asian			.189			.155
			(.035)			(.037)
Other/missing race			−.166			−.189
			(.118)			(.117)
High school top 10%			.067			.064
			(.020)			(.020)
High school rank missing			.003			−.008
			(.025)			(.023)
Athlete			.107			.092
			(.027)			(.024)
Average SAT score of schools applied to ÷ 100				.110	.082	.077
				(.024)	(.022)	(.012)
Sent two applications				.071	.062	.058
				(.013)	(.011)	(.010)
Sent three applications				.093	.079	.066
				(.021)	(.019)	(.017)
Sent four or more applications				.139	.127	.098
				(.024)	(.023)	(.020)

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column shows coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

TABLE 2.5
Private school effects: Omitted variables bias

	Dependent variable					
	Own SAT score ÷ 100			Log parental income		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	1.165	1.130	.066	.128	.138	.028
	(.196)	(.188)	(.112)	(.035)	(.037)	(.037)
Female		−.367			.016	
		(.076)			(.013)	
Black		−1.947			−.359	
		(.079)			(.019)	
Hispanic		−1.185			−.259	
		(.168)			(.050)	
Asian		−.014			−.060	
		(.116)			(.031)	
Other/missing race		−.521			−.082	
		(.293)			(.061)	
High school top 10%		.948			−.066	
		(.107)			(.011)	
High school rank missing		.556			−.030	
		(.102)			(.023)	
Athlete		−.318			.037	
		(.147)			(.016)	
Average SAT score of schools applied to ÷ 100			.777			.063
			(.058)			(.014)
Sent two applications			.252			.020
			(.077)			(.010)
Sent three applications			.375			.042
			(.106)			(.013)
Sent four or more applications			.330			.079
			(.093)			(.014)

Notes: This table describes the relationship between private school attendance and personal characteristics. Dependent variables are the respondent's SAT score (divided by 100) in columns (1)–(3) and log parental income in columns (4)–(6). Each column shows the coefficient from a regression of the dependent variable on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Omitted variables bias: example

Assessing omitted variables bias:

- $\hat{\beta}^s = 0.212$
- $\beta^s = \beta^\ell + \pi_1\gamma$
- What do you think the signs of π_1 and γ are?
- The estimated $\widehat{\pi_1} = 1.165$ (the difference in SAT scores between private and public college students) and $\hat{\gamma} = 0.051$
- So, $0.212 = \beta^\ell + (1.165 * 0.051)$. Our estimator of β using β_s is likely biased upward.
- $\hat{\beta}^\ell = 0.152$ (compare to column (2))

Example

Of course, a model with two explanatory variables is probably not sufficient in this example: it alone is unlikely to describe differences in average potential outcomes. Column (3) of Table 2.3 includes additional student covariates, such as log parental income, gender, race/ethnicity, athlete, and HS top 10%. The reduction in $\hat{\beta}$ suggests the estimator used in column (2) was still biased upward.

In a setting like this, one should still be concerned about *unobserved* omitted variables.

Example

In an attempt to address these, columns (4) - (6) represent what might be called a "self-revelation" model. They include the number and characteristics of schools to which students *applied*. This behavior might proxy for unobserved differences that are related to both private college attendance and earnings.

Example

Table 2.3
Private school effects: Average SAT score controls

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212	.152	.139	.034	.031	.037
	(.060)	(.057)	(.043)	(.062)	(.062)	(.039)
Own SAT score ÷ 100		.051	.024		.036	.009
		(.008)	(.006)		(.006)	(.006)
Log parental income			.181			.159
			(.026)			(.025)
Female			−.398			−.396
			(.012)			(.014)
Black			−.003			−.037
			(.031)			(.035)
Hispanic			.027			.001
			(.052)			(.054)
Asian			.189			.155
			(.035)			(.037)
Other/missing race			−.166			−.189
			(.118)			(.117)
High school top 10%			.067			.064
			(.020)			(.020)
High school rank missing			.003			−.008
			(.025)			(.023)
Athlete			.107			.092
			(.027)			(.024)
Average SAT score of schools applied to ÷ 100				.110	.082	.077
				(.024)	(.022)	(.012)
Sent two applications				.071	.062	.058
				(.013)	(.011)	(.010)
Sent three applications				.093	.079	.066
				(.021)	(.019)	(.017)
Sent four or more applications				.139	.127	.098
				(.024)	(.023)	(.020)

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column shows coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Table 2.5
Private school effects: Omitted variables bias

	Dependent variable					
	Own SAT score ÷ 100			Log parental income		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	1.165	1.130	.066	.128	.138	.028
	(.196)	(.188)	(.112)	(.035)	(.037)	(.037)
Female		−.367			.016	
		(.076)			(.013)	
Black		−1.947			−.359	
		(.079)			(.019)	
Hispanic		−1.185			−.259	
		(.168)			(.050)	
Asian		−.014			−.060	
		(.116)			(.031)	
Other/missing race		−.521			−.082	
		(.293)			(.061)	
High school top 10%		.948			−.066	
		(.107)			(.011)	
High school rank missing		.556			−.030	
		(.102)			(.023)	
Athlete		−.318			.037	
		(.147)			(.016)	
Average SAT score of schools applied to ÷ 100			.777			.063
			(.058)			(.014)
Sent two applications			.252			.020
			(.077)			(.010)
Sent three applications			.375			.042
			(.106)			(.013)
Sent four or more applications			.330			.079
			(.093)			(.014)

Notes: This table describes the relationship between private school attendance and personal characteristics. Dependent variables are the respondent's SAT score (divided by 100) in columns (1)–(3) and log parental income in columns (4)–(6). Each column shows the coefficient from a regression of the dependent variable on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Example

In columns (4) - (6) the estimated coefficient on private school shrinks and becomes statistically insignificant.

Interestingly, the correlation between *own* SAT score and private school enrollment is eliminated once application behavior has been controlled for (the self-revelation model). See column (3) of Table 2.5.

Ceteris paribus?

Even with rich controls we may remain concerned that the CEF we are estimating is not a description of how potential outcomes relate to our explanatory variable of interest. Example of Dinardo & Pischke (1997) on the returns to computer use on the job.

The techniques covered in this course are methods that have been developed to address this concern, in the absence of a randomized experiment.

Regression anatomy

The "regression anatomy" formula is a useful algebraic property of regression. Suppose X_1 is a causal variable of interest (e.g., private college attendance) and X_2 is a control (e.g., SAT score). Then:

$$\beta_1 = \frac{Cov(Y_i, \tilde{X}_{1i})}{V(\tilde{X}_{1i})}$$

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

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

Intuitively, "purge" X_{1i} of its covariance with X_{2i}, and regress Y_i on the residual.

Regression anatomy

This extends to models with more than 2 regressors:

$$\beta_K = \frac{Cov(Y_i, \tilde{X}_{Ki})}{V(\tilde{X}_{Ki})}$$

where $\tilde{X}_{Ki}$ is the *residual* from a regression of X_{Ki} on *all other* covariates.

Also known as the Frisch-Waugh-Lovell theorem.