Lecture 1: Single Equation OLS and IV

January 12, 2025

Linear Structural Equation

- The workhorse in empirical research of all kinds starts with a model linear in parameters.

- The model stated in terms of a (well-defined) population is:

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u = \beta_0 + x\beta + u,$$

  where x is $1 \times K$ and observed, and β is the $K \times 1$ vector of unknown "slope" parameters.

- Example:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{educ} \times \text{exper} + \beta_4 \text{exper}^2 + \beta_5 \text{female} + u$$

Linear Structural Equation

- In what follows, we assume that we can collect a random sample - that is, independent and identically distributed outcomes - from underlying population. Given randomly sampled observations $\{(\mathbf{x}_i, y_i) : i = 1, \ldots, N\}$ satisfying the population model before, we can write for the random draws:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i, i = 1, \ldots, N, \tag{1}$$

where N is the sample size.

Assumption 1 (OLS.1 Zero Correlation)

The error has a zero mean and is uncorrelated with each explanatory variables:

$$E(\mathbf{x}'u) = 0.$$

with the last equality a normalization (with an intercept in the model)

- Sufficient for assumption 1 is the stronger zero conditional mean assumption:

$$E(u|\mathbf{x}) = E(u) = 0.$$

Linear Structural Equation

Assumption 2 (OLS.2 No Perfect Collinearity)

In the population, there are no exact linear relationships among the covariates:

$$rank\ E(\mathbf{x}'\mathbf{x}) = \mathcal{K}$$

- The column (row) rank of a matrix A is the maximum number of linearly independent column (row) vectors of A.
- Violations: None in interesting application. High correlation among regressors often cannot be avoided, but not a violation of assumptions. Sometimes high correlation among regressors (multicollinearity) is the researcher's fault because parameterization has not been carefully chosen.
- When an intercept is included, 2 says the population variance-covariance matrix of the regressor is invertible.

Linear Structural Equation

Theorem 1 (Identification of OLS)

Under OLS.1 and OLS.2, β is identified, that is, we can write it as a function of population moments in observable variables:

$$
\begin{aligned}
\mathbf{x}'y &= (\mathbf{x}'\mathbf{x})\boldsymbol{\beta} + \mathbf{x}'u \\
E(\mathbf{x}'y) &= E(\mathbf{x}'\mathbf{x})\beta + E(\mathbf{x}'u) \\
E(\mathbf{x}'\mathbf{y}) &= E(\mathbf{x}'\mathbf{x})\boldsymbol{\beta} \;\; \text{By OLS.1} \\
\boldsymbol{\beta} &= [E(\mathbf{x}'\mathbf{x})]^{-1}E(\mathbf{x}'y) \;\; \text{By OLS.2}
\end{aligned}
$$

Linear Structural Equation

Theorem 2 (Estimation of OLS)

Apply the logic of method of moments given the random sample: replace population means with sample means:

$$\begin{aligned}\hat{\beta} &= \left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{y}_i\right)\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\end{aligned}$$

where $\mathbf{x}$ is $N \times K$ with i^{th} row $\mathbf{x}_i$ and $\mathbf{Y}$ is $N \times 1$ with i^{th} entry y_i

Theorem 3 (Consistency of OLS)

(Slutsky's Theorem of Plim on continuous functions): the probability limit of a continuous function is the value of that function evaluated at probability limit.

Linear Structural Equation

Theorem 4 (Asymptotic Normality)

To get the limiting distribution of OLS:

$$\sqrt{N}\left(\widehat{\beta}-\beta\right)=\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right) \tag{2}$$

Now, by the central limit theorem for i.i.d. random vectors,

$$N^{-1/2}\sum_{i=1}^{N}\mathbf{x}_i'u_i \xrightarrow{d} Normal(0,\mathbf{B}) \tag{3}$$

$$\mathbf{B}=Var(\mathbf{x}_i'u_i)=E(u_i^2\mathbf{x}_i'\mathbf{x}_i). \tag{4}$$

Linear Structural Equation

An implication of 3 is $N^{-1/2}\sum_{i=1}^{N} x_i' u_i = O_p(1)$. Given $\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1} - \mathbf{A}^{-1} = o_p(1)$, and using $O_p(1)\cdot o_p(1) = o_p(1)$.

$$\sqrt{N}\left(\hat{\beta}-\beta\right) = \mathbf{A}^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{x}_i' u_i\right) + o_p(1) \tag{5}$$

where $\mathbf{A} = E(\mathbf{x};\mathbf{x}_i)$ is $K\times K$ and nonsingular by OLS.2. Therefore, from 3 and 4:

$$\sqrt{N}\,(\hat{\beta}-\beta) \xrightarrow{d} Normal(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}). \tag{6}$$

The following Theorem is applied:

Theorem 5 (Lindeberg–Lévy CLT)

Suppose $(X_1, X_2, ...)$ is a sequence of i.i.d. random variables with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$. Then:

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0,\sigma^2).$$

Detour

Definition 1

- A sequence of random variables $\{x_N : N = 1, 2, 3, ...\}$ converges in probability to the constant a if for all $\varepsilon > 0$:

  $$\mathrm{P}\left[|x_N - a| > \varepsilon\right] \to 0 \qquad \text{as } N \to \infty$$

  We write $x_N \xrightarrow{p} a$ and say that a is the **probability limit (plim)** of x_N: $plim\ x_N = a$.

- In the special case where $a = 0$, we also say that $\{x_N\}$ is $o_p(1)$ (little oh p one). We also write $x_n = o_p(1)$ or $x_N \xrightarrow{p} 0$

- A sequence of random variables $\{x_N\}$ is **bounded in probability** if and only if for every $\varepsilon > 0$ there exists a $b_\varepsilon < \infty$ and an integer N_ε such that

  $$P[|x_N| > b_\varepsilon] < \varepsilon \text{ for all } N \geq N_\varepsilon$$

  We write $x_N = O_p(1)$ (big oh p one)

Linear Structural Equation

$$\hat{\mathbf{B}} = (N-K)^{-1} \sum_{i=1}^{N} \hat{u}_i^2 \mathbf{x}_i' \mathbf{x}_i \xrightarrow{p} \mathbf{B} \tag{7}$$

whether or not OLS.3 holds.
$Avar(\hat{\beta})$ is estimated with a "sandwich" form:

$$\begin{aligned} A\hat{v}ar(\hat{\beta}) &= \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N &(8)\\ &= \frac{N}{(N-K)} \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\sum_{i=1}^{N} \hat{u}_i^2 \mathbf{x}_i' \mathbf{x}_i \right) \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} &(9) \end{aligned}$$

- The factor $N/(N-K)$ is a finite-sample correction. Matters less for large N
- Note: R-squared is perfectly valid as a goodness-of-fit measure under heteroskedasticity. R^2 is a consistent estimate of $\rho^2 = 1 - \sigma_u^2/\sigma_y^2$, which is a function of the unconditional variances. Whether $Var(u|\mathbf{x})$ is constant is irrelevant for estimating ρ^2.

Potential Pitfalls of OLS

- Omitted variables bias/Selection
- Measurement error.
- Simultaneity

Instrumental Variables

- Start again with the population model:

$$y = \mathbf{x}\beta + u \tag{10}$$

where $\mathbf{x}$ is $1 \times K$, and in the vast majority of cases, $x_1 = 1$. This is the same model that we estimated by OLS.

- Let $\mathbf{z} = (z_1, z_2, ..., z_L)$ be a $1 \times L$ vector, where $z_1 = 1$ almost always. Further, $\mathbf{z}$ contains all exogenous elements of $\mathbf{x}$. But, if one or more elements of $\mathbf{x}$ is correlated with u, $\mathbf{z}$ must contain some outside variables.

- $\mathbf{z}$ is exogenous:

$$E(\mathbf{z}'u) = 0 \tag{11}$$

How does this assumption - these moment or orthogonality conditions - allow us to identify β?

Instrumental Variables

Theorem 6 (Identification with Instrumental Variables)

Suppose $L = K$. For example, $\mathbf{x} = (1, x_2, \ldots, x_{K-1}, x_K)$ and $z = (1, x_2, \ldots, x_{K-1}, z_1)$, so that only x_K is (possibly endogenous) and z_1 as an IV for x_K. Using 11 and 10:

$$\begin{aligned} E(\mathbf{z}'y) &= E(\mathbf{z}'\mathbf{x})\beta + E(\mathbf{z}'u) && (12) \\ &= E(\mathbf{z}'\mathbf{x})\beta \qquad \text{by condition 11} && (13) \end{aligned}$$

If we assume the *rank condition*:

$$rank\ E(\mathbf{z}'\mathbf{x}) = K \tag{14}$$

then

$$\beta = [E(\mathbf{z}'\mathbf{x})]^{-1} E(\mathbf{z}'y) \tag{15}$$

which extends the moment condition for OLS (which is the special case $\mathbf{z} = \mathbf{x}$)

Instrumental Variables

Given a random sample:

$$\hat{\beta}_{IV} = \left(N^{-1}\sum_{i=1}^{N} z_i' x_i\right)^{-1} \left(N^{-1}\sum_{i=1}^{N} z_i' y_i\right) \tag{16}$$

$$= \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{Y} \tag{17}$$

and $plim(\hat{\beta}_{IV}) = \beta$ under 11 and 14.

Testing for Rank Condition

- How can we test the rank condition? Difficult in general, but with a single endogenous explanatory variable, easy. Write the *reduced form* of x_K as
$$x_K = \delta_1 + \delta_2 x_2 + \ldots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_k$$
where, by definition,
$$E(r_k) = 0, \quad \mathrm{Cov}(x_j, r_k) = 0, \quad j = 2, \ldots, K-1, \quad \mathrm{Cov}(z_1, r_k) = 0.$$
- In other words, the linear projection of x_k on $(1, x_2, \ldots, x_{K-1}, z_1)$ is
$$L(x_K | 1, x_2, \ldots, x_{K-1}, z_1) = \delta_1 + \delta_2 x_2 + \ldots + \delta_{K-1} x_{K-1} + \theta_1 z_1.$$
- Can show: the rank condition 14 holds if and only if
$$\theta_1 \neq 0. \tag{18}$$
- OLS consistently estimates the parameters of a linear projection (not necessarily unbiased).
- Need to reject
$$H_0 : \theta_1 = 0$$
in favor of 18. Heteroskedasticity-robust inference can be used.

Multiple Instruments and 2SLS

- Again write

$$y = \mathbf{x}\beta + u$$

$$E(\mathbf{z}'u) = 0.$$

  where $L = \dim(\mathbf{z}) \geq \dim(\mathbf{x}) = K$

- In general, the best vector of IVs for x is the vector of linear projections of each element of x on z. We can write

$$\mathbf{x}_{1\times K} = \mathbf{z}_{1\times L}\Pi_{L\times K} + \mathbf{r}_{1\times K}$$

  where Π is the $L \times K$ matrix

$$\Pi = \left[E(\mathbf{z}'\mathbf{z})^{-1}\right]_{L\times L}\left[E(\mathbf{z}'\mathbf{x})\right]_{L\times K}$$

  and

$$E(\mathbf{z}'\mathbf{r}) = 0.$$

Multiple Instruments and 2SLS

- For each x_j we can write

$$x_j = \mathbf{z}\pi_j + r_j = x_j^* + r_j$$

where π_j $(L \times 1)$ is the jth column of Π.

- For any $x_j \in z, x_j^* = x_j$, so exogenous variables act as their own instruments.

- In the general case, use

$$\mathbf{x}^* = \mathbf{z}\Pi$$

as the $1 \times K$ vector of instruments for $\mathbf{x}$.

- Because z is exogenous, so is $\mathbf{x}^*$:

$$E(\mathbf{x}^{*\prime}u) = 0.$$

- The rank condition becomes

$$\text{rank } E(\mathbf{x}^{*\prime}\mathbf{x}) = K.$$

But

$$E(\mathbf{x}^{*\prime}\mathbf{x}) = \Pi' E(\mathbf{z}'\mathbf{x}) = E(\mathbf{x}'\mathbf{z})[E(\mathbf{z}'\mathbf{z})]^{-1}E(\mathbf{z}'\mathbf{x}).$$

Multiple Instruments and 2SLS

Deriving 2SLS: With

$$\beta = \left(E[\mathbf{x}^{*'}\mathbf{x}]\right)^{-1} E[\mathbf{x}^{*'}y],$$

need to worry about unknown Π because $\mathbf{x}_i^* = \mathbf{z}_i\Pi$. Two-step estimation:

1. Run the regression x_i on $z_i, i = 1, \ldots, N$ to obtain $\hat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$. Obtain the vector of fitted values,

   $$\hat{\mathbf{x}}_i = \mathbf{z}_i\hat{\Pi}, i = 1, \ldots, N.$$

   (This is the same as regressing each element of x_i not in z_i on z_i, and obtaining the fitted values. Any element of x_i in z_i is used as its own fitted value.)

2. Use $\hat{\mathbf{x}}_i$ as the vector of IVs for x_i:

   $$\hat{\beta}_{IV} = \left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{x}}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{x}}_i'y_i\right).$$

We can write this differently. Because

$$x_i = \hat{\mathbf{x}}_i + r_i \quad \sum_{i=1}^{N}\hat{\mathbf{x}}_i'r_i = 0 \quad \text{(by OLS FOCs)}.$$

Multiple Instruments and 2SLS: Estimation and Consistency

So

$$\sum_{i=1}^{N} \hat{\mathbf{x}}_i' \mathbf{x}_i = \sum_{i=1}^{N} \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i,$$

and then the IV estimator can be written as a two stage least squares estimator:

$$\hat{\beta}_{2SLS} = \left(N^{-1} \sum_{i=1}^{N} \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \hat{\mathbf{x}}_i' y_i \right).$$

Using full data matrices and some algebra, we can write:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \left(\hat{\mathbf{X}}' \hat{\mathbf{X}} \right)^{-1} \hat{\mathbf{X}}' \mathbf{Y} \\
&= \left[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X}) \right]^{-1} (\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{Y}) \\
&= \beta + \left[(\mathbf{X}'\mathbf{Z}/N)(\mathbf{Z}'\mathbf{Z}/N)^{-1}(\mathbf{Z}'\mathbf{X}/N) \right]^{-1} (\mathbf{X}'\mathbf{Z}/N)(\mathbf{Z}'\mathbf{Z}/N)^{-1}(\mathbf{Z}'\mathbf{U}/N)
\end{aligned}$$

where the last expression can be used to show consistency by applying the WLLN to each term, along with the rank condition and $E(\mathbf{z}'u) = 0$

Multiple Instruments and 2SLS: Asymptotic

$$N^{1/2}(\hat{\beta}_{2SLS} - \beta) = \left(N^{-1} \sum_{i=1}^{N} (\hat{\mathbf{x}}_i)' \hat{\mathbf{x}}_i \right)^{-1} \left(N^{-1/2} \sum_{i=1}^{N} (\hat{\mathbf{x}}_i)' u_i \right) + o_p(1)$$

where the $\mathbf{x}_i^* = z_i \Pi$ are the linear projections.

$$N^{1/2}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$$

$$\mathbf{A} = E(\mathbf{x}_i^{*\prime}\mathbf{x}_i^*), \quad \mathbf{B} = E(u_i^2 \mathbf{x}_i^{*\prime}\mathbf{x}_i^*)$$

Consistent (not unbiased) estimators of σ^2 and $\mathbf{A}$:

$$\hat{\sigma}^2 = (N-K)^{-1} \sum_{i=1}^{N} \hat{u}_i^2 \xrightarrow{p} \sigma^2, \quad \hat{\mathbf{A}} = N^{-1} \sum_{i=1}^{N} \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i$$

Heteroskedasticity-Robust Inference:

$$\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N = \frac{N}{N-K} \left(\sum_{i=1}^{N} \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i \right)^{-1} \left(\sum_{i=1}^{N} \hat{u}_i^2 \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i \right) \left(\sum_{i=1}^{N} \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i \right)^{-1}$$

Empirical Example: Card (1995)

- Return to schooling – casual correlation shows earning gains of 5-15 percent per additional year of schooling.
- Key confounding factor – education levels are endogenous choices, may over-state or under-state "true" return
- Data: young men cohort of national longitudinal survey
- Key idea: use geographical proximity of nearby 4-year college to local labor market
- IV estimate of return to education 25-60 % higher tan OLS estimate (OLS understates the return to schooling)

Empirical Example: Card (1995)

Data:

- National Longitudinal Survey of Young Men: began in 1966, through 1981
- Non-white residents over-sample: high fraction of southern region, black
- Characteristics of local labor market, including an indicator for the presence of an accredited 4-year college in the local labor market
- Labor market information from 1976 interviews, less sample attrition
- Left with 3010 observations with wage records.

Empirical Example: Card (1995) Regression Results

	(1)	(2)	(3)	(4)	(5)
1. Education	0.074 (0.004)	0.075 (0.003)	0.073 (0.004)	0.074 (0.004)	0.073 (0.004)
2. Experience	0.084 (0.007)	0.085 (0.007)	0.085 (0.007)	0.085 (0.007)	0.085 (0.007)
3. Experience-Squared /100	-0.224 (0.032)	-0.229 (0.032)	-0.230 (0.032)	-0.226 (0.032)	-0.229 (0.032)
4. Black Indicator	-0.190 (0.017)	-0.199 (0.018)	-0.194 (0.019)	-0.194 (0.019)	-0.189 (0.019)
5. Live in South	-0.125 (0.015)	-0.148 (0.026)	-0.146 (0.026)	-0.145 (0.026)	-0.146 (0.026)
6. Live in SMSA	0.161 (0.015)	0.136 (0.020)	0.136 (0.020)	0.137 (0.020)	0.138 (0.020)
7. Region in 1966 (8 indicators)	no	yes	yes	yes	yes
8. Live in SMSA in 1966	no	yes	yes	yes	yes
9. Parental Education[a] (main effects)	no	no	yes	yes	yes
10. Interacted Parental Education Classes[b]	no	no	no	yes	yes
11. Family Structure[c] (2 indicators)	no	no	no	no	yes

Figure: Card (1995) Regression Table

Empirical Example: Card (1995) Economic Model

Consider a two-equation system describing schooling (S_i) and log wages (y_i) for individual i (in 1976):

$$S_i = X_i \gamma + v_i, \tag{1}$$
$$y_i = X_i \alpha + S_i \beta + u_i. \tag{2}$$

Here X_i is a vector of observed attributes (with $E(X_i u_i) = E(X_i v_i) = 0$) and β has the interpretation of the "true" return to education. A conventional earnings equation estimated by OLS suffers from the correlation between u and v:

- Ability bias (say IQ score): upward bias
- Measurement error in **schooling**: downward bias

Empirical Example: Card (1995)

Argument by Card (1995):

> *(...) presence of a nearby college may be such a variable. Students who grow up in an area without a college face a higher cost of college education, since the option of living at home is precluded. One would expect this higher cost to reduce investments in higher education, at least among children from relatively low-income families.*

Validation of partial correlation:

- Regress schooling on region, urban/rural indicator (1966), age, and race dummies, family structure, parental education.

Empirical Example: Card (1995)

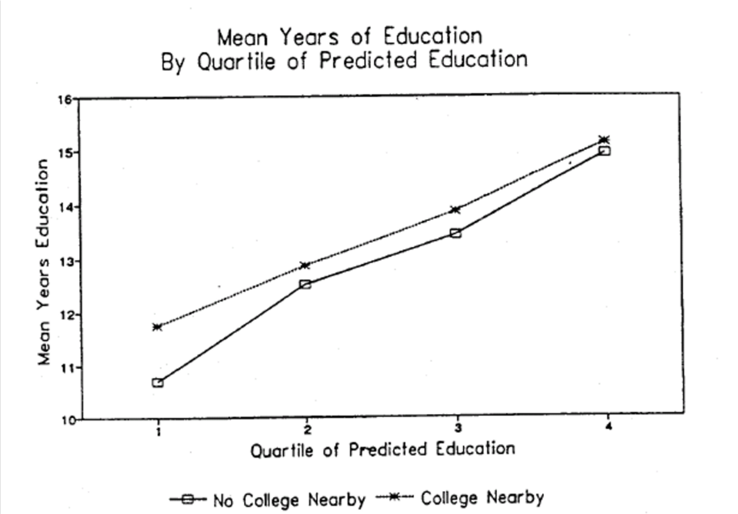

Figure: Card(1995) Figure

Empirical Example: Card (1995)

Table 3: Reduced Form and Structural Estimates of Education and Earnings Models

	Reduced Form Models: Education		Reduced Form Models: Earnings		Structural Models of Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
A: Treat Experience and Experience Squared as Exogenous						
1. Live Near College in 1966	0.320 (0.088)	0.322 (0.083)	0.042 (0.018)	0.045 (0.018)	--	--
2. Education	--	--	--	--	0.132 (0.055)	0.140 (0.055)
3. Family Background Variables[a]	no	yes	no	yes	no	yes
B: Treat Experience and Experience Squared as Endogenous[b/]						
4. Live Near College in 1966	0.382 (0.114)	0.365 (0.105)	0.047 (0.019)	0.048 (0.019)	--	--
5. Education	--	--	--	--	0.122 (0.046)	0.132 (0.049)
6. Family Background Variables[a]	no	yes	no	yes	no	yes

Figure: Card(1995) Table 3

Empirical Example: Card (1995) Discussions

2SLS estimate is much larger than OLS

- But confidence interval (95%) much wider too.

What if experience is also mis-measured?

- Paper uses age, age-squared as additional instruments.

Is classic measurement error a good explanation?
Why proximity to college can also be endogenous

- "high ability" family chooses to live closer to college.
- Impact of college on elementary/secondary school quality
- Unobserved geographic wage premium

These explanations are "direct earning effect" of proximity to college.
Additional data variation based on family income.

Empirical Example: Card (1995) Discussions

The interpretation of college proximity as a factor that lowers the cost of higher education suggests that growing up near a college should have a bigger effect on the education outcomes of children from poorer families. The pattern of education differentials in Figure 1 confirms this notion. Letting X_{1i} denote the components of X_i **other than** college proximity, the implied model for schooling is:

$$S_i = X_{1i}\gamma_1 + C_i\delta_0 + C_i \times P_i\delta_1 + v_i, \tag{1b}$$

where C_i is an indicator for growing up near a college, P_i is an indicator for low family income, and the coefficients δ_0 and δ_1 are both positive. In this case, even if C_i is included directly in the earnings equation:

$$y_i = X_{1i}\alpha_1 + C_i\alpha_0 + S_i\beta + u_i, \tag{2b}$$

the interaction $C_i \times P_i$ of college proximity and poor family background can be used as an instrumental variable for education.

Empirical Example: Card (1995) Discussions

Table 5: Instrumental Variables Estimates of the Return to Education Based on Interaction of Parental Education and Proximity to College

	Reduced Form Models: Education	Reduced Form Models: Earnings	Structural Models of Earnings	
	(1)	(2)	(3)	(4)
1. Live Near College in 1966	0.154 (0.135)	0.029 (0.024)	0.015 (0.029)	0.013 (0.024)
2. Live Near College * Low Parental Education[a]	0.462 (0.186)	0.043 (0.032)	--	--
3. Education[b]	--	--	0.093 (0.065)	0.097 (0.048)
4. Family Background Variables[c]	yes	yes	yes	yes

Figure: Card(1995) Table 5