Part I

Contents

References

- Wooldridge (2013) chapter 3
- Stock and Watson (2009) chapter 6 (parts of 7,8,9, & 18 also relevant)
- Angrist and Pischke (2014) chapter 2
- Bierens (2010)
- Baltagi (2002) chapter 4
- Linton (2017) Part III (more advanced)

1 Introduction

Why we need a multiple regression model

- There are many factors affecting the outcome variable y.
- If we want to estimate the marginal effect of one of the factors (regressors), we need to control for other factors.
- Suppose that we are interested in the effect of x_1 on y , but y is affected by both x_1 and x_2:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i \tag{1}$$

Why we need a multiple regression model

- Assume:

MLR.1 (linear model)

MLR.2 (independence) $\{(x_{1,i}, x_{2,i}, y_i)\}_{i=1}^n$ is independent random sample

MLR.3 (rank condition) no multicollinearity

MLR.4 (exogeneity) $\mathrm{E}[\epsilon|X] = 0$

- Suppose we regress y only against x_1 :

$$\hat{\beta}_1^s = \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1) y_i}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2}$$

- What is $\mathrm{E}[\hat{\beta}_1^s]$?

Why we need a multiple regression model

- What is $\mathrm{E}[\hat{\beta}_1^s]$?

$$\begin{aligned}
\hat{\beta}_1^s =& \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i)}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} \\
=& \beta_1 + \beta_2 \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1) x_{2,i}}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} + \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1) \epsilon_i}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} \\
\mathrm{E}[\hat{\beta}_1^s | x\text{'s}] =& \beta_1 + \beta_2 \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1) x_{2,i}}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} \\
=& \beta_1 + \beta_2 \frac{\widehat{\mathrm{Cov}}(x_1, x_2)}{\widehat{\mathrm{Var}}(x_1)}
\end{aligned}$$

- $\hat{\beta}_1^s$ biased unless $\widehat{\text{Cov}(x_1, x_2)} = 0$ or $\beta_2 = 0$

We will call the regression of y on only x_1 the short regression, and the regression of y on x_1 and x_2 the long regression. The short regression coefficient on x_1, β_1^s captures how as x_1 varies there is both a direct impact on y, β_1, and an indirect impact through how x_2 varies with x_1, $\beta_2 \frac{\widehat{\text{Cov}}(x_1,x_2)}{\widehat{\text{Var}}(x_1)}$. When we look at the long regression, we eliminate this indirect channel, effectively holding x_2 constant. This is a very important concept. In multivariate regression, the interpretation of the slope coefficients is that they are the effect of each x variable holding all others constant. For example, when there are two regressors, β_1 is the effect of x_1 holding x_2 constant.

Omitted variable bias

- When true model is "long regression"

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$$

  but we only estimate the "short regression",

$$y_i = \beta_0 + \beta_1^s x_{1,i} + \underbrace{v_i}_{=\beta_2 x_{2,i} + \epsilon_i}$$

  then

$$\text{E}[\hat{\beta}_1^s] = \beta_1 + \beta_2 \frac{\widehat{\text{Cov}}(x_1, x_2)}{\widehat{\text{Var}}(x_1)}$$

- If x_1 and x_2 related, we can no longer say that $\text{E}[v_i | x_{1,i}] = 0$
- When x_1 changes, x_2 changes as well, which contaminates estimation of the effect of x_1 on y
- As a result, the short regression estimate, $\hat{\beta}_1^s$, is biased

---

Omitted variable bias is a useful way for thinking about how a violation of the exogeneity assumption might affect our estimates. For example, consider a regression of wages on years of eduction. Let $y =$ log wage, $x =$ years of education. The regression model is

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where we want β_1 to be the causal effect of education on wages. It is likely that there is some unobserved ability, say a_i that effects both wages and education. Since ability effects wages, a_i is part of ϵ_i. Thus, it seems unlikely that $\text{E}[\epsilon_i | x_i] = 0$, so OLS is likely biased. If we think about omitted variables bias, we can be more specific about the bias. Decompose ϵ as

$$\epsilon_i = \beta_2 a_i + u_i$$

where u_i represents all unobservables that affect wages and are uncorrelated with ability. Then we have

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 a_i + u_i.$$

Omitted variables bias tells us that

$$\mathrm{E}[\hat{\beta}_1^s] = \beta_1 + \beta_2 \frac{\widehat{\mathrm{Cov}}(x,a)}{\widehat{\mathrm{Var}}(x)}.$$

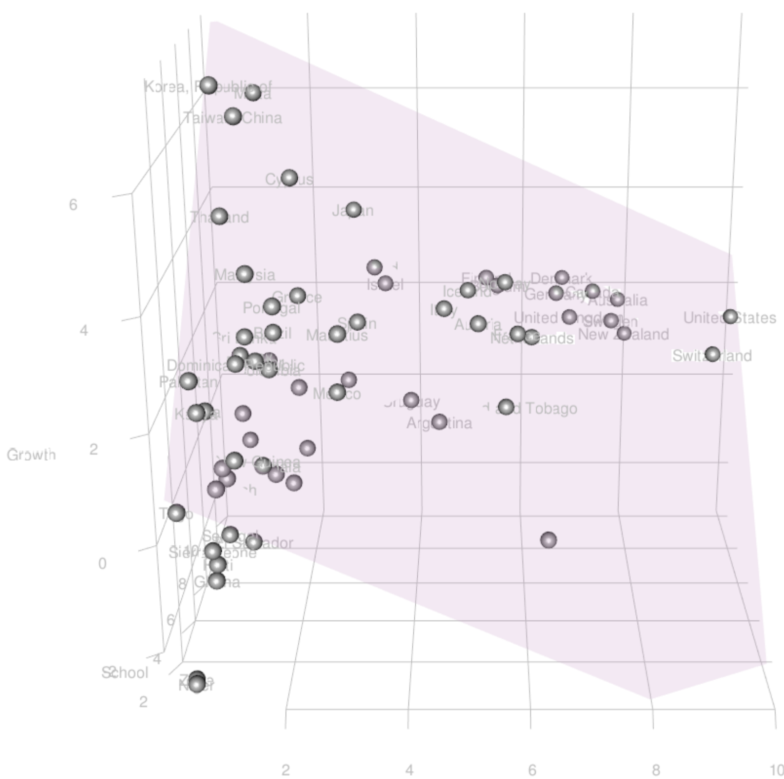
We expect ability to be positively related to wages ($\beta_2 > 0$) and education ($\mathrm{Cov}(x,a) > 0$), thus we expect the bivariate regression of log wages on education to produce an upward biased estimate of the causal effect of education on wages, i.e.

$$\mathrm{E}[\hat{\beta}_1^s] = \beta_1 + \beta_2 \frac{\widehat{\mathrm{Cov}}(x,a)}{\widehat{\mathrm{Var}}(x)} > \beta_1.$$

2 Examples

2.1 Example: growth, GDP, and schooling

Example: growth, GDP, and schooling



	Model 1	Model 2	Model 3
(Intercept)	1.796***	0.958*	0.895*
	(0.378)	(0.418)	(0.389)
rgdp60	0.047		−0.485**
	(0.095)		(0.146)
yearsschool		0.247**	0.640***
		(0.089)	(0.144)
R^2	0.004	0.110	0.244
Adj. R^2	-0.012	0.095	0.219
Num. obs.	65	65	65
RMSE	1.908	1.804	1.676

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Growth and GDP and education in 1960

Example: growth, GDP, and schooling

2.2 California test scores

Test scores and student teacher ratios

- Data from Stock and Watson (2009)
- School average test score and student teacher ratios
- Code

Test scores and student teacher ratios

2.3 Chandra et al. (2008)

---

[0] Code

	1	2	3	4
β_0	698.93***	739.45***	638.73***	640.32***
	(9.47)	(60.72)	(7.45)	(5.77)
$\frac{student}{teacher}$	−2.28***	−6.45	−0.65*	−0.07
	(0.48)	(6.19)	(0.35)	(0.28)
$\left(\frac{student}{teacher}\right)^2$		0.11		
		(0.16)		
Income			1.84***	1.49***
			(0.09)	(0.07)
English learners				−0.49***
				(0.03)
Num. obs.	420	420	420	420

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2: Test scores and student teacher ratios

Chandra et al. (2008)

"Does Watching Sex on Television Predict Teen Pregnancy?"

- Long regression:

$$preg = \beta_0 + \beta_1 sexTV + \beta_2 totalTV + \epsilon$$

  - $preg$ = teen pregnancy
  - $sexTV$ = hours of television with sexual content
  - $totalTV$ = hour of television

- Short (bivariate) regression:

$$preg = \beta_0^s + \beta_1^s sexTV + v$$

- Results:
  - Short: $\hat{\beta}_1^s \sqrt{\widehat{\text{Var}}(sexTV)} = 0.18$
  - Long: $\hat{\beta}_1 \sqrt{\widehat{\text{Var}}(sexTV)} = 0.30$
  - Also report that

$$totalTV = \gamma_0 + \gamma_1 sexTV + e$$

$\hat{\gamma}_1 \sqrt{\widehat{\text{Var}}(sexTV)} = 0.29$

- What is β_2?
  - From omitted variable bias formula, we know

$$\begin{aligned} \text{E}[\hat{\beta}_1^s] =& \beta_1 + \beta_2 \underbrace{\frac{\widehat{\text{Cov}}(sexTV, totalTV)}{\widehat{\text{Var}}(sexTV)}}_{=\hat{\gamma}_1} \\ =& \beta_1 + \beta_2 \hat{\gamma}_1 \\ \beta_2 =& \frac{\beta_1 - \text{E}[\hat{\beta}_1^s]}{\hat{\gamma}_1} \end{aligned}$$

– so,

$$\begin{aligned}\hat{\hat{\beta}}_2 =& \frac{\hat{\beta}_1 - \hat{\beta}_1^s}{\hat{\gamma}_1} \\ =& \frac{(\hat{\beta}_1 - \hat{\beta}_1^s)\sqrt{\widehat{\text{Var}}(sexTV)}}{\hat{\gamma}_1\sqrt{\widehat{\text{Var}}(sexTV)}} \\ =& \frac{0.30 - 0.18}{0.29} \\ =& -0.41\end{aligned}$$

---

3 Interpretation

Multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$

As in bivariate regression, let's assume:

MLR.1 (linear model)

MLR.2 (independence) $\{(x_{1,i}, x_{2,i}, y_i)\}_{i=1}^n$ is independent random sample

MLR.3 (rank condition) no multicollinearity

MLR.4 (exogeneity) $\text{E}[\epsilon | x_{1,i}, x_{2,i}, \dots, x_{k,i}] = 0$

Interpretation of coefficients

- β_j is partial (marginal) effect of x_j on y

$$\beta_j = \frac{\partial y_i}{\partial x_{j,i}}$$

- β_j is partial (marginal) effect of $\text{E}[y|x_1, \dots, x_k]$

$$\beta_j = \frac{\partial \text{E}[y|x_1, \dots, x_k]}{\partial x_j}$$

- Other regressors are held constant

$$\Delta y_i = \beta_0 + \beta_1 \Delta x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$

β_1 is the *ceteris paribus* effect of $x_{1,i}$ on y_i

Changing more than one regressor simultaneously

- Sometimes it does not make sense to change x_j without also changing x_k
- Example: age-earnings profile

$$\log wage_i = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \epsilon_i$$

  - Cannot age while holding age^2 constant
  - Instead should report marginal effect

$$\frac{\partial \log wage_i}{\partial age_i}|_{age_i = a} = \beta_1 + 2\beta_2 a$$

  or average marginal effect

$$\overline{\frac{\partial \log wage_i}{\partial age_i}} = \beta_1 + 2\beta_2 \overline{age}$$

Changing more than one regressor simultaneously

- Example: Chandra et al. (2008) effect of exposure to sexual content on television on teen pregnancy

$$preg = \beta_0 + \beta_1 sexTV + \beta_2 totalTV + \epsilon$$

  - $sexTV$ = hours of television with sexual content
  - $totalTV$ = hour of television
  - Should we care about β_1, β_2, or some combination?

TABLE 1 Bivariate Analyses Predicting Pregnancy After Baseline Among Youths Who Ever Had Sex ($N = 718$)

Baseline Predictors	β	P
Exposure to sex on television[a]	0.18	.280
Covariates		
Total television exposure[a]	−0.23	.162
Age	0.20	.108
Lower grades	0.05	.749
Parent education[a]	−0.16	.223
Educational aspirations[a]	−0.13	.405
Hispanic[b]	0.04	.923
Black[b]	0.95[c]	.011
Female	0.99[d]	.002
Lives in 2-parent household	−1.50[d]	<.001
Deviant or problem behavior[a]	0.42[c]	.012
Intention to have children before age 22 y	1.03[c]	.013

[a] Coefficients reflect the association with pregnancy for each 1-SD increase or decrease in the predictor.

[b] Comparison group is all other races.

[c] $P < .05$.

[d] $P < .01$.

TABLE 2 Multivariate Logistic Regression Analyses Predicting Pregnancy After Baseline Among Youths Who Ever Had Sex ($N = 718$)

Baseline Predictors	β	P
Exposure to sex on television[a]	0.44[b]	.034
Covariates		
Total television exposure[a]	−0.42[b]	.022
Age	0.28[b]	.022
Lower grades	0.21	.288
Parent education[a]	0.00	.999
Educational aspirations[a]	−0.14	.446
Hispanic[c]	0.86	.084
Black[c]	1.20[b]	.011
Female	1.20[d]	.001
Lives in 2-parent household	−1.50[d]	<.001
Deviant or problem behavior[a]	0.43[b]	.014
Intention to have children before age 22	0.61	.279

[a] Coefficients reflect the association with pregnancy for each 1-SD increase or decrease in the predictor.
[b] $P < .05$.
[c] Comparison group is white and races other than Hispanic or black.
[d] $P < .01$.

4 OLS estimation

OLS estimation

$$(\hat{\beta}_0, \dots, \hat{\beta}_k) = \underset{b_0,\dots,b_k}{\arg\min} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1,i} - \cdots - b_k x_{k,i})^2$$

- First order conditions:

$$\begin{aligned}
\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i} \right) &= 0 \\
\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i} \right) x_{1,i} &= 0 \\
\vdots &= \vdots \\
\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i} \right) x_{k,i} &= 0
\end{aligned}$$

or with $\hat{\epsilon}_i = \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \cdots - \hat{\beta}_k x_{k,i} \right)$,

$$\begin{aligned}
\sum_{i=1}^{n} \hat{\epsilon}_i =& 0 \\
\sum_{i=1}^{n} \hat{\epsilon}_i x_{j,i} =& 0 \text{ for } j = 1, 2, ..., k
\end{aligned}$$

- We choose $\hat{\beta}_0, ..., \hat{\beta}_k$ so that residuals and regressors are uncorrelated (orthogonal)
- First order conditions are a system of linear equations in $\hat{\beta}_0, ..., \hat{\beta}_k$

The first order conditions are a system of linear equations in $\hat{\beta}_0, ..., \hat{\beta}_k$. It is possible to write down an explicit algebraic expression for the $\hat{\beta}_j$, but it would be quite complicated and not very useful.[1]

5 Partitioned regression

Although a completely explicit solution to the first order conditions is complicated, it is possible to write a recursive formula for $\hat{\beta}_j$. This approach is called partitioned regression or partialling out. It will be useful in proofs and for thinking about bias in multiple regression.

Partitioned regression

- A useful representation of $\hat{\beta}_j$ (e.g. $j = 1$)
  - Regress $x_{1,i}$ on other regressors

    $$x_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2,i} + \cdots + \hat{\gamma}_k x_{k,i} + \tilde{x}_{1,i}$$

    where $\tilde{x}_{1,i}$ is the OLS residual

[1] If you have taken linear algebra and are comfortable working with matrices, then you can write the OLS coefficients in a not too complicated way. (If you are not familiar with matrices, then you can safely ignore this footnote). Let X be an $n \times (k+1)$ matrix where the first column are all ones and the ith, jth for entry for $j > 1$ is $x_{j-1,i}$. Let Y be an $n \times 1$ vector consisting of the y_i, and let $\hat{\beta}$ be $(k+1) \times 1$. Then the first order conditions can be written as $X'(Y - X\hat{\beta}) = 0$. Solving for $\hat{\beta}$ we have $\hat{\beta} = (X'X)^{-1}(X'Y)$.

– Then

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{x}_{1,i} y_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2}$$

– Note that $\frac{1}{n}\sum_{i=1}^n \tilde{x}_{1,i} = 0$, so $\hat{\beta}_1$ = slope coefficient from regressing y on $\tilde{x}_1$

Partitioned regression gives a recursive formula for $\hat{\beta}_1$ in that it involves the residuals from the regression of $x_{1,i}$ on the other x's. This is itself a multiple regression. However, it is a multiple regression with one less independent variable. We could apply partitioned regression again to calculate the $\hat{\gamma}$'s. At some point the first step of partitioned regression would just involve a single independent variable and we'd be done. For example, with $k = 3$, we can say

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{x}_{1,i} y_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2}$$

where $\tilde{x}_{1,i}$ are the residuals from

$$x_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2,i} + \hat{\gamma}_3 x_{3,i} + \tilde{x}_{1,i}.$$

Now, using partitioned regression again, we can say that

$$\hat{\gamma}_2 = \frac{\sum_{i=1}^n \tilde{\tilde{x}}_{2,i} x_{1,i}}{\sum_{i=1}^n \tilde{\tilde{x}}_{2,i}^2}$$

where $\tilde{\tilde{x}}_{2,i}$ are residuals from

$$x_{2,i} = \hat{\delta}_0 + \hat{\delta}_3 x_{3,i} + \tilde{\tilde{x}}_{2,i}$$

and

$$\hat{\delta}_3 = \frac{\sum_{i=1}^n (x_{3,i} - \bar{x}_3) y_i}{\sum_{i=1}^n (x_{3,i} - \bar{x}_3)^2}.$$

In practice, we rarely need to write down or work with all the steps back in the recursion, but it is important to understand that partitioned regression completely describes how to calculate the $\hat{\beta}_j$.

To show that partitioned regression works, we need to show that it gives the same $\hat{\beta}_1$ as the OLS first order conditions.

Proof outline

- Substitute $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \cdots + \hat{\beta}_k x_{k,i} + \hat{\epsilon}_i$ into $\frac{\sum_{i=1}^n \tilde{x}_{1,i} y_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2}$
- Use the following facts to simplify:
  1. $\sum_{i=1}^n \tilde{x}_{1,i} = 0$
  2. $\sum_{i=1}^n \tilde{x}_{1,i} x_{j,i} = 0$ for $j = 2, ..., k$
  3. $\sum_{i=1}^n \tilde{x}_{1,i} x_{1,i} = \sum_{i=1}^n \tilde{x}_{1,i}^2$
  4. $\sum_{i=1}^n \tilde{x}_{1,i} \hat{\epsilon}_i = 0$
- See handout version of slides for details

Proof. To begin with let's assume the 4 facts are true. If we substitute $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \cdots + \hat{\beta}_k x_{k,i} + \hat{\epsilon}_i$ into the partitioned regression formula we have

$$\begin{aligned}
\frac{\sum_{i=1}^n \tilde{x}_{1,i} y_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2} =& \frac{\sum_{i=1}^n \tilde{x}_{1,i} \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \cdots + \hat{\beta}_k x_{k,i} + \hat{\epsilon}_i\right)}{\sum_{i=1}^n \tilde{x}_{1,i}^2} \\
& \text{distributing and rearranging} \\
=& \hat{\beta}_0 \frac{\sum_{i=1}^n \tilde{x}_{1,i}}{\sum_{i=1}^n \tilde{x}_{1,i}^2} + \hat{\beta}_1 \frac{\sum_{i=1}^n \tilde{x}_{1,i} x_{1,i}}{\sum_{i=1}^n \tilde{x}_{1,i}^2} + \hat{\beta}_2 \frac{\sum_{i=1}^n \tilde{x}_{1,i} x_{2,i}}{\sum_{i=1}^n \tilde{x}_{1,i}^2} + \cdots + \hat{\beta}_k \frac{\sum_{i=1}^n \tilde{x}_{1,i} x_{k,i}}{\sum_{i=1}^n \tilde{x}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{x}_{1,i} \hat{\epsilon}_i}{\sum_{i=1}^n \tilde{x}_{1,i}^2} \\
& \text{using the 4 facts} \\
=& \hat{\beta}_1
\end{aligned}$$

Now, we just need to show the 4 facts.

The $\tilde{x}_{1,i}$ are residuals from

$$x_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2,i} + \cdots + \hat{\gamma}_k x_{k,i} + \tilde{x}_{1,i}.$$

The first order conditions for this regression are

$$\begin{aligned}
\sum_{i=1}^n \tilde{x}_{1,i} =& 0 \\
\sum_{i=1}^n \tilde{x}_{1,i} x_{2,i} =& 0 \\
\vdots =& \vdots \\
\sum_{i=1}^n \tilde{x}_{1,i} x_{k,i} =& 0.
\end{aligned}$$

This is exactly facts 1 and 2.

Fact 3 comes from

$$\begin{aligned}
\sum_{i=1}^n \tilde{x}_{1,i} x_{1,i} =& \sum_{i=1}^n \tilde{x}_{1,i} \left(\hat{\gamma}_0 + \hat{\gamma}_2 x_{2,i} + \cdots + \hat{\gamma}_k x_{k,i} + \tilde{x}_{1,i}\right) \\
=& \hat{\gamma}_0 \sum_{i=1}^n \tilde{x}_{1,i} + \hat{\gamma}_2 \sum_{i=1}^n \tilde{x}_{1,i} x_{2,i} + \cdots + \hat{\gamma}_k \sum_{i=1}^n \tilde{x}_{1,i} x_{k,i} + \sum_{i=1}^n \tilde{x}_{1,i} \tilde{x}_{1,i} \\
& \text{using facts 1 and 2} \\
=& \sum_{i=1}^n \tilde{x}_{1,i}^2.
\end{aligned}$$

Finally, since $\hat{\epsilon}_i$ are residuals from regressing y on the x's, we have

$$0 = \sum_{i=1}^n \hat{\epsilon}_i = \sum_{i=1}^n \hat{\epsilon}_i x_{1,i} = \cdots = \sum_{i=1}^n \hat{\epsilon}_i x_{k,i}.$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{n} \tilde{x}_{1,i}\hat{\epsilon}_i =& \sum_{i=1}^{n} \left(x_{1,i} - \hat{\gamma}_0 - \hat{\gamma}_2 x_{2,i} - \cdots - \hat{\gamma}_k x_{k,i} \right) \hat{\epsilon}_i \\ =& \sum_{i=1}^{n} x_{1,i}\hat{\epsilon}_i - \hat{\gamma}_0 \sum_{i=1}^{n} \hat{\epsilon}_i - \hat{\gamma}_2 \sum_{i=1}^{n} \hat{\epsilon}_i x_{2,i} - \cdots - \hat{\gamma}_k \sum_{i=1}^{n} \hat{\epsilon}_i x_{k,i} \\ =& 0,\end{aligned}$$

which shows fact 4. □

"Partialling out"

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{x}_{1,i} y_i}{\sum_{i=1}^{n} \tilde{x}_{1,i}^2}$$

1. Regress x_1 against the other controls (regressors) and keep the residuals $\tilde{x}_1$, which is the "part" of x_1 that is uncorrelated with the other regressors
2. Regress y against $\tilde{x}_1$

$\hat{\beta}_1$ measures the effect of x_1 after the effects of $x_2, \ldots, x_k$ have been partialled out

---

$\tilde{x}_{1,i}$ is the part of $x_{1,i}$ that is uncorrelated with the other regressors. This is another way of seeing that multiple regression gives the relationship between x_1 and y holding the other x's constant. The multiple regression slope coefficient,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{x}_{1,i} y_i}{\sum_{i=1}^{n} \tilde{x}_{1,i}^2},$$

looks at the covariance of y with the part of $x_{1,i}$ that is uncorrelated with the other x's. In contrast, the short regression coefficient

$$\hat{\beta}_1^s = \frac{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1) y_i}{\sum_{i=1}^{n} (x_{1,i} - \bar{x}_1)^2},$$

looks at the covariance of y with all of $x_{1,i}$, some of which may be related to the other x's.

5.1 Example: California test scores

Example: California test scores and student teacher ratios

```
## long regression
summary(lm(testscr ~ str + avginc + calw_pct + meal_pct + el_pct, data=ca))
## partialling out
tilde.str <- residuals(lm(str ~ avginc + calw_pct + meal_pct + el_pct,
                          data=ca))
mean(tilde.str) ## should be 0
cov(tilde.str, ca$avginc) ## should be 0
sum(tilde.str*ca$str)
```

```
sum(tilde.str^2) ## should equal previous line

## should equal long regression coefficient
cov(tilde.str, ca$testscr)/var(tilde.str)
sum(tilde.str*ca$testscr)/sum(tilde.str^2)

## =long regression coefficient, but wrong standard error
summary(lm(ca$testscr ~ tilde.str))
```

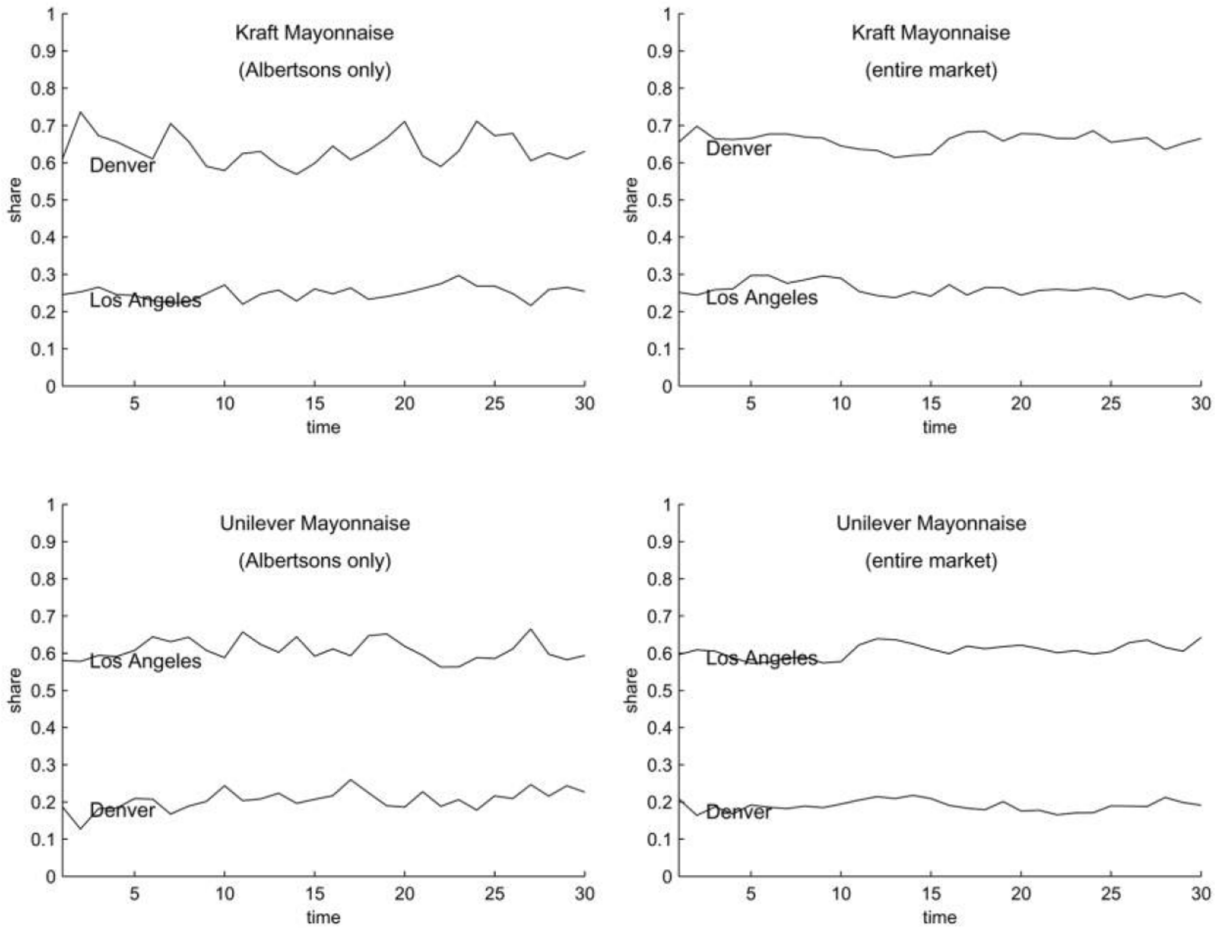
---

6 Example: Bronnenberg, Dhar, and Dubé (2009)

Example: Bronnenberg, Dhar, and Dubé (2009)

- Looks at market shares of brands of consumer packaged goods (CPG) across markets and time
  - CPG = beer, coffee, ketchup, etc.
- Market shares from AC Nielsen scanner data
  - This type of data has been used very frequently in IO during the last decade
  - AC Nielsen distributes bar code scanners to a sample of consumers, consumers record every purchase by scanning bar codes
  - 4-week intervals, June 1992-May 1995
- Results
  - Market shares variable across geographic markets, but persistent over time within each market
  - Geographic market shares strongly correlated with first mover advantage
    - e.g. Miller (founded in Milwaukee) most popular beer in Milwaukee, Budweiser (founded in St. Louis) most popular beer in St. Louis

---



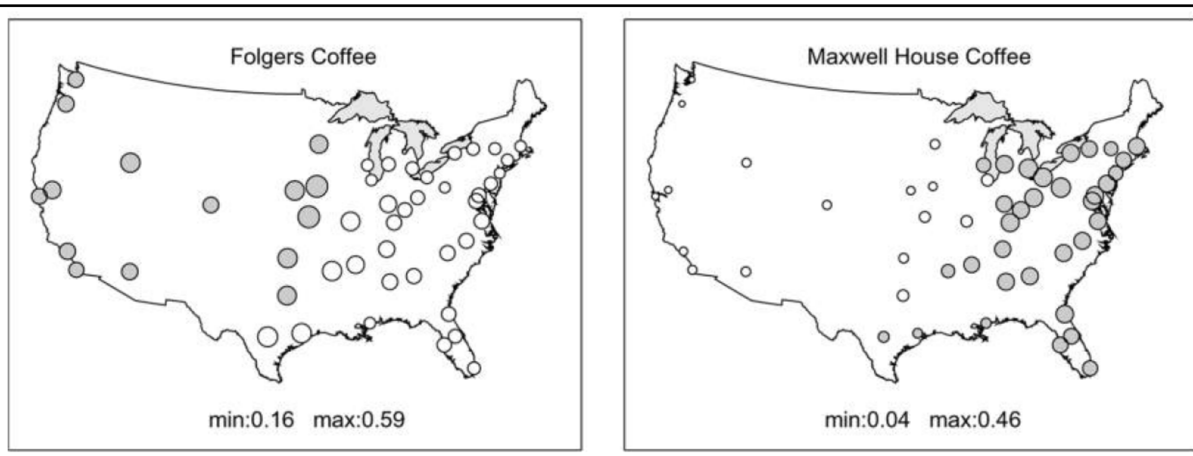


FIG. 2.—The joint geographic distribution of share levels and early entry across U.S. markets in ground coffee. The areas of the circles are proportional to share levels. Shaded circles indicate that a brand locally moved first.

$$share_{im} = \beta_0 + \beta_1 \underbrace{brandA_{im}}_{=1 \text{ if good } i \text{ is brand A}} + \beta_2 \underbrace{earlyEntry_{im}}_{=1 \text{ if good } i \text{ entered } m \text{ first}} + \epsilon_{im}$$

Variable	Entry Effect (1)	Brand Effects (2)	Entry and Brand Effects (3)
Beer ($N = 94$):			
Intercept	.141 (.010)	.149 (.011)	.139 (.011)
Budweiser		.118 (.016)	.020 (.026)
Miller			
Early entry	.134 (.014)		.117 (.026)
R^2	.483	.372	.487
Coffee ($N = 150$):			
Intercept	.139 (.011)	.059 (.014)	.052 (.011)
Folgers		.251 (.020)	.206 (.015)
Maxwell House		.197 (.020)	.088 (.018)
Hills Bros.			
Early entry	.208 (.019)		.175 (.015)
R^2	.440	.533	.755
Ketchup ($N = 50$):			
Intercept			.388 (.019)
Heinz			
Early entry			.072 (.025)
R^2			.149

- How are results in the three columns for beer related?

$$\hat{\beta}_1^s = \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1) y_i}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2}$$

- Substitute $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\epsilon}_i$

$$
\begin{aligned}
\hat{\beta}_1^s =& \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\epsilon}_i)}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} \\
=& \frac{\hat{\beta}_0 \overbrace{\left(\sum_{i=1}^n x_{1,i} - \bar{x}_1 \right)}^{=0} + \hat{\beta}_1 \left(\sum_{i=1}^n (x_{1,i} - \bar{x}_1) x_{1,i} \right)}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} + \\
&+ \frac{\hat{\beta}_2 \left(\sum_{i=1}^n (x_{1,i} - \bar{x}_1) x_{2,i} \right)}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} + \\
&+ \frac{\overbrace{\left(\sum_{i=1}^n x_{1,i} \hat{\epsilon}_i \right)}^{=0 \text{FOC}} - \overbrace{\left(\bar{x}_1 \sum_{i=1}^n \hat{\epsilon}_{1,i} \right)}^{=0 \text{FOC}}}{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2}
\end{aligned}
$$

-

$$
\begin{aligned}
\hat{\beta}_1^s =& \hat{\beta}_1 + \hat{\beta}_2 \frac{\widehat{\text{Cov}}(x_1, x_2)}{\widehat{\text{Var}}(x_1)} \\
=& \hat{\beta}_1 + \hat{\beta}_2 \hat{\gamma}_1^2
\end{aligned}
$$

  where $\hat{\gamma}_1^2$ is coefficient on x_1 in a regression of x_2 on x_1

- "Short ($\hat{\beta}_1^s$) equals long ($\hat{\beta}_1$) plus the effect of omitted ($\hat{\beta}_2$) times the regression of the omitted on the included ($\hat{\gamma}_1^2$)" (Angrist and Pischke, 2009)

- Similarly,

$$
\hat{\beta}_2^s = \hat{\beta}_2 + \hat{\beta}_1 \underbrace{\hat{\gamma}_2^1}_{= \frac{\widehat{\text{Cov}}(x_1, x_2)}{\widehat{\text{Var}}(x_1)}}
$$

This formula is nearly identical to omitted variables bias. Omitted variables bias is about the $\text{E}[\hat{\beta}_1^s]$ and involves population regression coefficients. It is that

$$
\text{E}[\hat{\beta}_2^s | X] = \beta_2 + \beta_1 \widehat{\text{Cov}}(x_1, x_2) \widehat{\text{Var}}(x_1).
$$

We just the similar fact that

$$
\hat{\beta}_2^s = \hat{\beta}_2 + \hat{\beta}_1 \underbrace{\hat{\gamma}_2^1}_{= \frac{\widehat{\text{Cov}}(x_1, x_2)}{\widehat{\text{Var}}(x_1)}}.
$$

The former follows from the later by taking expectations conditional on X and assuming our usual four regression assumptions so that $\text{E}[\hat{\beta}_j | X] = \beta_j$.

- In this example, $\hat{\beta}^s_{bud} = 0.118$, $\hat{\beta}^s_{entry} = 0.134$, $\hat{\beta}_{bud} = 0.020$, and $\hat{\beta}_{entry} = 0.117$, so

$$\begin{aligned}\frac{\widehat{\text{Cov}}(bud, entry)}{\widehat{\text{Var}}(entry)} &= \frac{\hat{\beta}^s_{entry} - \hat{\beta}_{entry}}{\hat{\beta}_{bud}} \\ &= \frac{0.134 - 0.117}{0.02} = 0.85 \\ \frac{\widehat{\text{Cov}}(bud, entry)}{\widehat{\text{Var}}(bud)} &= \frac{\hat{\beta}^s_{bud} - \hat{\beta}_{bud}}{\hat{\beta}_{entry}} \\ &= \frac{0.118 - 0.02}{0.17} = 0.84\end{aligned}$$

References

References

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Baltagi, BH. 2002. *Econometrics*. Springer, New York. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001086635&title=Econometrics.

Bierens, Herman J. 2010. "Multivariate Linear Regression." URL http://personal.psu.edu/hxb11/LINREG3.PDF.

Bronnenberg, B.J., S.K. Dhar, and J.P.H. Dubé. 2009. "Brand history, geography, and the persistence of brand shares." *Journal of Political Economy* 117 (1):87–115. URL http://www.jstor.org/stable/10.1086/597301.

Chandra, A., S.C. Martino, R.L. Collins, M.N. Elliott, S.H. Berry, D.E. Kanouse, and A. Miu. 2008. "Does watching sex on television predict teen pregnancy? Findings from a national longitudinal survey of youth." *Pediatrics* 122 (5):1047–1054. URL http://www.pediatricsdigest.mobi/content/122/5/1047.short.

Linton, Oliver B. 2017. *Probability, Statistics and Econometrics*. Academic Press. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001868500&title=Probability%2C%20statistics%20and%20econometrics.

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.
