Session 3: OLS (II)

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

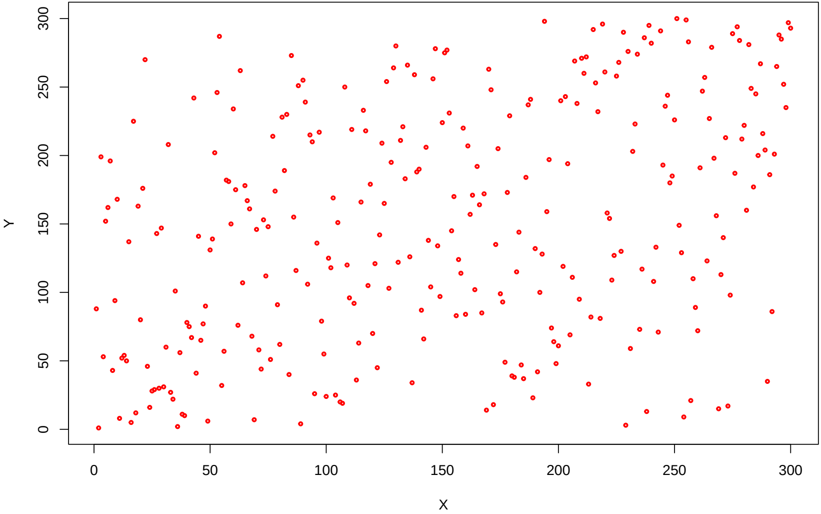
Last time...

- Descriptive statistics: first, second moment
- Association: covariance, correlation
- Linear model (theory)
- Estimation of bivariate LM: ordinary least squares (OLS)

Plan for today

- Multiple regression
- Readings: Stock and Watson (2011) (ch6), Verbeek (2018) (ch2.8-2.9, 3)
- Also: Cameron and Trivedi (2005) (ch4.1-4.5), Greene (2008) (ch2-3)

Guessing game...



Correlation is...?

Regression: Advanced

- So far: single treatment D + estimate (maybe) of ATE, $\hat{\beta}$
- Recall: $\hat{\beta}_{ols} = Cov(D,Y)/V(D)$.
- I said that OLS *can* generate unbiased, consistent, efficient estimates
- Two problems
  - Conditions under which $\hat{\beta}$ meets desirable criteria?
  - How to estimate models with additional variables?

Notation

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,a} & z_{1,b} & \cdots & w_{1,k} \\ 1 & x_{2,a} & z_{2,b} & \cdots & w_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,a} & z_{n,b} & \cdots & w_{n,k} \end{bmatrix} \quad \beta = \begin{bmatrix} \alpha_0 \\ \beta_1 \\ \gamma_2 \\ \vdots \\ \tau_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

With n observations and $k+1$ independent variables (the 'regular' independent variables plus a constant/y-intercept):

For unit i:

$$
\begin{aligned}
y_i &= \begin{bmatrix} 1 & x_{i,a} & z_{i,b} & \cdots w_{i,k} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_1 \\ \gamma_2 \\ \vdots \\ \\ \tau_k \end{bmatrix} + u_i \\
&= \mathbf{x}_i' \beta + \varepsilon_i
\end{aligned}
$$

Compactly:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

where $\mathbf{X}$ are stacked $\mathbf{x}'$ vectors.

- Same optimization (minimization) problem as before...
- Let's solve it!

Done?

Problem to be solved:

$$\min \beta \text{SSR} \equiv \mathbf{u}'\mathbf{u}$$

Assumptions: no perfect multicollinearity, existence of moments

Unique solution:

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \left(\sum_i^n X_i X_i' \right)^{-1} \sum_i^n X_i Y_i \end{aligned}$$

To find this solution:

$$\begin{aligned}
\mathbf{u}'\mathbf{u} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\
&= (\mathbf{y}' - \hat{\beta}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\beta}) && \text{(by properties of transpose)} \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} - \hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} && \text{(distributing)} \\
&= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} && \text{(reorganizing)}
\end{aligned}$$

We can then solve the minimization problem:

$$
\begin{aligned}
\frac{\partial \mathsf{SSR}}{\partial \hat{\beta}} &= 0 - 2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0 \\
2\mathbf{X}'\mathbf{X}\hat{\beta} &= 2\mathbf{X}'\mathbf{y} \\
\mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{X}'\mathbf{y} \\
(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
\mathbf{I}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}
\end{aligned}
$$

Example

Consider the following example with k=2 independent variables, x_1 and x_2, and $n = 5$ observations.

Observation	x_1	x_2	y
1	4	11	27
2	2	5	18
3	1	13	24
4	6	7	31
5	8	15	27

Example

The matrices look like this (the column of 1's in $\mathbf{X}$ being the future constant/y-intercept):

$$\mathbf{X} = \begin{bmatrix} 1 & 4 & 11 \\ 1 & 2 & 5 \\ 1 & 1 & 13 \\ 1 & 6 & 7 \\ 1 & 8 & 15 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 27 \\ 18 \\ 24 \\ 31 \\ 27 \end{bmatrix}$$

Example

We then want to compute two quantities: $(\mathbf{X}'\mathbf{X})^{-1}$ (assuming it exists) and $\mathbf{X}'\mathbf{y}$.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & 6 & 8 \\ 11 & 5 & 13 & 7 & 15 \end{bmatrix} \begin{bmatrix} 1 & 4 & 11 \\ 1 & 2 & 5 \\ 1 & 1 & 13 \\ 1 & 6 & 7 \\ 1 & 8 & 15 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 21 & 51 \\ 21 & 121 & 222 \\ 51 & 222 & 654 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} \approx \begin{bmatrix} 2.4946 & -0.2973 & -0.1096 \\ -0.2973 & 0.2432 & -0.0811 \\ -0.1096 & -0.0811 & 0.0270 \end{bmatrix}$$

Example

Then:

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & 6 & 8 \\ 11 & 5 & 13 & 7 & 15 \end{bmatrix} \begin{bmatrix} 27 \\ 18 \\ 24 \\ 31 \\ 27 \end{bmatrix}$$

$$= \begin{bmatrix} 127 \\ 546 \\ 1422 \end{bmatrix}$$

Example

We can now put everything together:

$$\hat{\beta} = \begin{bmatrix} 2.4946 & -0.2973 & -0.1096 \\ -0.2973 & 0.2432 & -0.0811 \\ -0.1096 & -0.0811 & 0.0270 \end{bmatrix} \begin{bmatrix} 127 \\ 546 \\ 1422 \end{bmatrix}$$

$$\hat{\beta} \approx \begin{bmatrix} -3.7432 \\ 4.7027 \\ 0.6216 \end{bmatrix}$$

Or:

$$y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$$
$$\hat{y} = -3.7 + 4.7x_1 + 0.6x_2$$

Interpretation

- We often say: $\hat{\beta}$ is the effect of D **holding Z,... constant**
- Or better: $\hat{\beta}$ is the effect of D **adjusting for Z,...**
- Recall: $\partial Y / \partial D = \beta$. Thus, $\hat{\beta}$ is the estimated effect of a small change in D (adjusting for Z)
- But what does this mean? What does OLS do to our data?
- Two ways: Frisch and Waugh (1933) and Morgan and Winship (2014). For now: **Frisch-Waugh Theorem** aka **regression anatomy theorem**. (See also Angrist and Pischke (2008), Filoso (2013).)

Frisch-Waugh Theorem

(1) Suppose your model is:

$$y_i = \alpha + \beta d_i + \delta w_i + ... + \gamma z_{i,k} + \varepsilon_i$$

(2) Define the residual $\tilde{d}_i = d_i - \hat{d}_i$ from:

$$d_i = \rho + \lambda w_i + ... + \tau z_i + \mu_i$$
$$\hat{d}_i = \hat{\rho} + \hat{\lambda} w_i + ... + \hat{\tau} z_i$$

(3) Then, we can express $\hat{\beta}$ as:

$$\hat{\beta} = \frac{Cov(Y, \tilde{d})}{V(\tilde{d})}$$

Implications of Frisch-Waugh

- With multiple regressors: $\hat{\beta}$ is the effect of D on Y without variation in D that can be explained by the other variables
- Variable W will only affect $\hat{\beta}$ if W is correlated with D. To see this: $\hat{d}_i$ will vary if $\hat{\lambda} \neq 0$ and W is/isn't included.
- We can guess why **randomization** of D will help: in expectation, it makes D uncorrelated with any other variable. Thus, we won't have to think (too much) about adding variables or not.

Conclusion

Summary so far...

1. We started from the problem: how can we estimate the causal effect of D on Y
   - Eg effect of an ad campaign on sales, of a new product on revenues, on a carbon tax on CO2, etc.
2. **Fundamental problem of causal inference** shows why we can never recover the treatment effect. We have to **infer** it.
3. To organize thoughts, we built a linear model (theory)
4. We showed how OLS estimates the parameter(s) of the model
5. We did so with very few assumptions so far: no perfect multicollinearity
6. Via Frisch-Waugh, we clarified what we are doing to the data

Next questions

- How set up our models
- How confident can we be about our estimates of the treatment effect?
- How good is our model?
- Later: what can we deal with assumption failures

Questions?

References

Angrist, Joshua, and Steffan J. Pischke. 2008. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. London: Cambridge University Press.

Filoso, Valerio. 2013. "Regression Anatomy, Revealed." *Stata Journal* 13 (1): 92–106.

Frisch, Ragnar, and Frederick V. Waugh. 1933. "Partial Time Regressions as Compared with Individual Trends." *Econometrica* 1 (4): 387–401.

Greene, William H. 2008. *Econometric Analysis*. 6th ed. New York: Pearson.

Morgan, Steven L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd Edition. Cambridge: Cambridge University Press.

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.