Lecture 8: Empirical Methods in Corporate Finance I

Adam Hal Spencer

The University of Nottingham

Advanced Financial Economics 2019

Roadmap

1. Introduction
2. Basic Regression Model
3. Forms of Endogeneity
4. Potential Outcomes
5. Conclusion

Motivation

- The theory section gave us some qualitative insights into how the five financial frictions affect firm value, investment behaviour and capital structure.
- In the remainder of the corporate finance part of this course, we ask the question of, which frictions matter the most in the real world?

Data

- There are three broad classifications of data.

(1) Cross-section ($i \in \{1, 2, ..., N\}$) [*fixed time*].

(2) Time series: ($t \in \{1, 2, ..., T\}$) [*fixed variable*].

(3) Panel: (*it* with $i \in \{1, 2, ..., N\}$ and $t \in \{1, 2, ..., T\}$).

Roadmap

1. Introduction
2. Basic Regression Model
3. Forms of Endogeneity
4. Potential Outcomes
5. Conclusion

Linear regression

- A simple cross-sectional (population) linear regression model typically takes the form

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_M x_{M,i} + u_i \quad (1)$$

  where

  - y_i is the outcome or dependent variable.
  - $\{x_{1,i}, x_{2,i}, \ldots, x_{M,i}\}$ are the explanatory variables.
  - $\{\beta_0, \beta_1, \ldots, \beta_M\}$ are the coefficient parameters to be estimated.
  - u_i is an unobservable random error or disturbance term.

- The objective is to get estimates of the regression coefficients.

- We would like to be able to say, "an increase in $x_{1,i}$ of 1 unit leads to an increase in y_i of β_1 units."

Linear regression

- One can think of the terms involving the regression coefficients as a "model".
- I mean the word model here in the reduced-form sense, (c.f. the structural sense). More on this later.
- It's designed to be the movements in the dependent variable that are captured by changes in the explanatory variables.
- u_i you can think of as everything exogenous to our model.

Ordinary least squares (OLS)

- The most common way to estimate a regression equation is to use OLS.
- This estimator finds the coefficients that minimise the sum of squared residuals.
- Intuitively: minimises the sample equivalent squared "u_i" term in the regression specification (1).

Ordinary least squares (OLS)

- Re-write equation (1) in vector form

$$y_i = \beta x_i + u_i$$

where β and x_i are now vectors containing all the individual terms.

- Define the sum of squared residuals (SSR) as

$$\Omega(\beta) = \sum_{i=1}^{N} (y_i - \beta x_i)^2.$$

- The OLS estimator $\hat{\beta}$ is defined as

$$\hat{\beta} = \min_{\beta} \Omega(\beta)$$

- OLS chooses the coefficients to minimise the distance of the observed data from the regression model.

Ordinary least squares (OLS)

- The expression for the solution is given by

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

in vector notation. For the simpler version of $y_i = \beta_0 + \beta_1 x_i$ for just one regressor, we get

$$\hat{\beta}_1 = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Goodness of fit

- Once we have our estimates, we can assess the fit of the linear model.
- Define fitted values of the dependent variable as $\hat{y}_i = \hat{\beta} x_i$.
- The values of the dependent variable, which are predicted by the OLS estimates given the explanatory variable values.
- Coefficient of determination (R-squared) is a measure of goodness of fit, (how close does $\hat{y}_i$ get to y_i)?
- Defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

where y_i is the observed value, $\hat{y}_i$ is the fitted value and $\bar{y}$ is the sample mean.

Consistency of OLS

- A consistent estimator is such that estimates of a population parameter converge to the truth for asymptotically large samples.
- Denote this by $\hat{\beta}_i \to \beta_i$.
- OLS estimates are consistent provided that the following assumptions hold

(1) The sample is random,

(2) The error term is zero in expectation,

(3) There are no linear relationships between the explanatory variables,

(4) The error term is uncorrelated with the explanatory variables.

Endogeneity

- Asymptotic consistency is a good thing: means we're getting pretty close to the truth with the estimates.
- Endogeneity is when the error term is correlated with the explanatory variables, (i.e. assumption (4) fails).
- With endogeneity, our OLS estimates are no longer consistent!

What does endogeneity mean for corporate finance?

- A corporate finance researcher may be interested in a regression of the form

$$\text{Leverage}_i = \beta_0 + \beta_1 \text{Profitability} + u_i$$

where Leverage_i is the debt to equity ratio of the firm and Profitability is their net profits.

- Want to estimate: a 1 unit increase in profitability leads to a β_1 unit increase in leverage.
- Do we think that this is exogenous, (i.e. all good, the opposite of endogeneity).
- If we ran this regression in the presence of endogeneity, what would β_1 mean?

Roadmap

1. Introduction
2. Basic Regression Model
3. Forms of Endogeneity
4. Potential Outcomes
5. Conclusion

Forms

- Endogeneity comes in a variety of forms.

(i) Omitted variables,

(ii) Simultaneity,

(iii) Measurement error.

Omitted variables

- Probably the most obvious case.
- Say that the true economic relation is given by

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \gamma w_i + u_i$$

but we don't see anything about the variable w_i. So we run

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

where now $v_i = \gamma w_i + u_i$.

Omitted variables

- If w_i is uncorrelated with $x_{1,i}$ and $x_{2,i}$, then we're good.
- Say that w_i is correlated with $x_{1,i}$ but uncorrelated with $x_{2,i}$.
- Rare that this will be the case in corporate finance, but if it were, then $\hat{\beta}_2$ would still be consistent, (i.e. $\hat{\beta}_2 \rightarrow \beta_2$).
- However in the limit $\hat{\beta}_1 \rightarrow \beta_1 + \gamma \frac{\text{cov}(x_j, w)}{\text{Var}(x_j)}$.
- I don't expect you to prove this: just understand the following intuition.
- The asymptotic bias is made up of the effect of the omitted variable on the dependent variable γ and the effect of the independent variable $\frac{\text{cov}(x_j, w)}{\text{Var}(x_j)}$.

Omitted variables: in corporate finance

- What could w_i be in corporate finance?
- Information asymmetry: what is this?
- Very abstract concept. It can't be accurately measured.
- How is it correlated with the independent variables?
- What does it mean for regression inference?

Simultaneity

- Does x cause y or is it the other way around?
- Also referred to as reverse causation.

Simultaneity: in corporate finance

- Regress the firms' market to book ratio on an index of anti-takeover provisions.
- Negative regression coefficient.
- Can we interpret this as: an increase in anti-takeover provisions leads to a loss in this value ratio?
- Or could it be that managers of low value firms adopt stronger anti-takeover provisions to entrench themselves?
- Correlation v.s. causation.

Simultaneity bias: example

- Say that variables x and y are determined simultaneously via the following system

$$y = \beta x + u$$
$$x = \alpha y + v$$

where u is uncorrelated with v.

- Think of y as the market-book ratio and x as anti-takeover provisions.

Simultaneity bias: example

- Say we just regress y on x, (ignore the second equation).
- See that the OLS regression coefficient will be

$$\begin{aligned}\hat{\beta} &= \frac{\text{Cov}(x, y)}{\text{Var}(x)} \\ &= \frac{\text{Cov}(x, \beta x + u)}{\text{Var}(x)} \\ &= \beta + \frac{\text{Var}(x, u)}{\text{Var}(x)}.\end{aligned}$$

....biased!

Measurement error

- Say that we only see a noisy proxy for the dependent variable.
- True regression model is $y_i^* = \beta x_i + v_i$ where y_i^* is the true dependent variable.
- We only observe $y_i = y_i^* + u_i$.
- See then that a regression of

$$y_i = \beta x_i + w_i$$

where $w_i = u_i + v_i$ lends itself to the same issues as omitted variable bias.

Measurement error

- What about measurement error in the independent variables?
- If we assume the error term is uncorrelated with all explanatory variables, we're good.
- If not, our estimates are again biased.
- Can lead to all of our coefficient estimates being biased, even if the error is only correlated with one explanatory variable.

Roadmap

1. Introduction
2. Basic Regression Model
3. Forms of Endogeneity
4. Potential Outcomes
5. Conclusion

Treatment Effects

- Say there are two groups of firms that are susceptible to a treatment. Use the following notation
  - Y_{1i}: outcome for firm i if exposed to treatment ($D_i = 1$).
  - Y_{0i}: outcome for the same individual if not exposed ($D_i = 0$).
- These two outcomes are referred to as potential outcomes since we only observe the following in the data

$$Y_i = Y_{1i}D_i + Y_{i0}(1 - D_i)$$

that is — we don't observe the counterfactual for any given individual.

Treatment Effects

- We want to understand the causal effect of the treatment.
- We'd figure that out by holding everything constant and looking at how a given individual is affected.
- Missing data: we don't see the counterfactual.

Treatment Effects

- Ok so we need to estimate average effects.
- Define the following two objects
  - Average treatment effect (ATE): $\alpha_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}]$: the expected treatment effect of a subject randomly drawn from the population.
  - Average treatment effect on the treated (ATT): $\alpha_{ATT} \equiv \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$: the expected treatment effect for a firm that has been treated.

Treatment Effects

- A standard measure for estimating the treatment effect is to estimate parameter

$$
\begin{aligned}
\beta &\equiv \mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] \\
&= \mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] \\
&= \{\mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 1]\} + \{\mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_{0i} = 0]\}
\end{aligned}
$$

where the second line comes from adding and subtracting $\mathbb{E}[Y_{0i}|D_i = 1]$.

- What are these objects?

- Difference $\mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 1]$ is looking at the expected difference in the treated and untreated outcomes for an individual who has received the treatment.

Treatment Effects

- That is, we can decompose the estimator into

$$\beta = \alpha_{ATT} + B$$

where $\alpha_{ATT} \equiv \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$ from before and B is a selection bias term, given by

$$B = \mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_{0i} = 0]$$

Treatment Effects

- This bias term gives the difference in untreated outcomes for those who have been treated and have not been treated.
- A non-zero difference can stem from the situation where treatment status is the result of individual decisions where those with low Y_0 choose treatment more frequently than those with high Y_0.

What does this mean for corporate finance?

- How did the Tax Cuts and Jobs Act (TCJA) in early 2018 affect the investment behaviour of firms?

Roadmap

1 Introduction

2 Basic Regression Model

3 Forms of Endogeneity

4 Potential Outcomes

5 Conclusion

Summary

- My best friend in graduate school was an econometric theorist.
- He said "econometrics is all about trying to change one thing without changing another".
- This is the hard thing about data in social science: we can't run controlled experiments.
- It's hard to change an x variable without also changing something in the residual term u in economics.
- Experimental sciences can do this: have a controlled environment in a lab where only one thing is changed.

Summary

- OLS is a simple and powerful tool under the right assumptions.
- The big benefit is that we don't impose much structure on the relationship between the ys and xs.
- Hell can break loose in the face of endogeneity though. What fixes are there?