Lecture 2: Empirical Methods I

Adam Hal Spencer

The University of Nottingham

Advanced Financial Economics 2020

Roadmap

Motivation

- What's the point of theory?
- Different philosophies on this.
- One way to view it: to guide what we'd expect to see in the "real world".
- The real world is data.
- How can we use statistical/econometric techniques to map from economic ideas to raw data?

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

Linear regression

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

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ... + \beta_M x_{M,i} + u_i \tag{1}$$

  where

  - y_i is the outcome or dependent variable.
  - $\{x_{1,i}, x_{2,i}, ..., x_{M,i}\}$ are the explanatory variables.
  - $\{\beta_0, \beta_1, ..., \beta_M\}$ 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 (2).

Ordinary least squares (OLS)

- Re-write equation (2) 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

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

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

Forms

- Endogeneity comes in a variety of forms.

(i) Omitted variables,

(ii) Simultaneity,

(iii) Measurement error.

- We'll focus on (i) here.

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?

Roadmap

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

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.
- Luckily, we have some tricks up our sleeves.

Roadmap

General setting

- A common way to deal with endogeneity.
- Say that we're considering a framework of

$$y_i = \beta_0 + \beta_1 x_{1,i} + ... + \beta_M x_{M,i} + u_i \tag{2}$$

where $Cov(x_{k,i}, u_i) \neq 0$ for some $k \in \{1, 2, ..., M\}$.

- Generally means that all of our coefficient estimates will be biased.
- Unless $x_{k,i}$ happens to be uncorrelated with the other regressors.

General setting

- An instrumental variable (denoted z_i) for $x_{k,i}$ satisfies two conditions
  - **(1)** Relevance condition.
  - **(2)** Exclusion condition.

Relevance Condition

- Requires that the partial correlation between the instrument and endogenous variable not be zero.
- Means formally that the coefficient γ in the regression

$$\begin{aligned} x_{k,i} =& \alpha_0 + \alpha_1 x_{1,i} + ... + \alpha_{k-1} x_{k-1,i} + \alpha_{k+1} x_{k+1,i} \\ &+ ... + \alpha_M x_{M,i} + \gamma z_i + v_i \end{aligned}$$

  be non-zero.
- Says that the endogenous variable and the instrument are correlated after netting-out the effects of all other exogenous variables.

Exclusion Condition

- Says that z_i's only influence on the outcome variable of interest is through the endogenous regressor.
- That is: $Cov(z_i, u_i) = 0$ where u_i is the residual from (2) above.
- What does it mean if this condition is not satisfied?
- Would mean that the instrument is also endogenous! This is the same problem that we're trying to fix.

Multiple variables

- One can use multiple instruments for an endogenous variable.
- Both conditions need to be satisfied for each instrument though.
- Relevance condition can be done with joint test of statistical significance.
- May also have multiple endogenous variables.
- In this case, need at least as many instruments as we have endogenous variables.

Examples in corporate finance

- Bennedsen et al. (2007): does replacing an outgoing CEO with a family member hurt firm performance in family firms?
- CEO succession is likely correlated with things that affect performance also.
- E.g. non-family CEO may come in during bad times, family CEO during good times.
- Need exogenous variation in the CEO succession decision.
- They use gender of the first-born child of a departing CEO.
- Show that CEOs with boy-first families are significantly more likely to appoint a family CEO.
- Gender is a biological thing, likely uncorrelated with the firm's performance.

Estimation

- Common approach is to use two-stage least squares (2SLS).

- **1st stage:** regress the endogenous variable ($x_{k,i}$) on the exogenous variables and instruments. Gives fitted values $\hat{x}_{k,i}$: those predicted by the regression.

$$\hat{x}_{k,i} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1,i} + \ldots + \hat{\alpha}_{k-1} x_{k-1,i} + \hat{\alpha}_{k+1} x_{k+1,i} + \ldots + \hat{\alpha}_M x_{M,i} + \hat{\gamma} z_i$$

- **2nd stage:** use the fitted values ($\hat{x}_{k,i}$) to stand-in for the endogenous variable ($x_{k,i}$) in the regression of y_i (outcome) on all the right-side variables.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_k \hat{x}_{k,i} + \ldots \hat{\beta}_M x_{M,i}$$

Estimation

- The residual in the first stage contains all the junk that's correlated with the outcome variable.
- We disregard that and just keep the exogenous component, (since recall z_i is doesn't affect the outcome variable directly).
- The idea is that the fitted values $\hat{x}_{k,i}$ contain only variation in the endogenous variable that is exogenous to the regression system.
- We've removed the "evil" endogenous part.
- The regression coefficient will be consistent.
- Introduces more noise into the estimation though: need to correct the standard error estimates.

Estimation

- Easy to run two stage least squares in Stata.
- If we're dealing with a model

$$y_i = \beta x_i + u_i$$

where z_i is an instrument for x_i then use the Stata command

```
ivregress 2sls y (x = z)
```

Roadmap

Panel setup

- A panel dataset for firms follows the same firm for multiple periods of time.
- Are there unobservable attributes at the firm-level that are time-invariant?
- We could control for these with fixed effects.

Fixed effects

- A fixed effects panel regression takes the form

$$y_{it} = \beta x_{it} + \alpha_i + u_{it}$$

  where parameter α_i is a firm-specific, unobservable and time-invariant fixed effect.

- We can't observe this fixed effect.

- I.e. it's likely an omitted variable (absorbed into the residual term).

- That is: there might be omitted variables that are correlated with the firm itself.

- So how is it helpful conceptually?

Least Squares with Dummy Variables

- Why don't we just use a dummy for each firm (each i)?
- Run the following regression

$$y_{it} = \beta x_{it} + \sum_{j=1}^{N} \alpha_j d_j + u_{it}$$

where $d_j = 1$ if $j = i$ and 0 otherwise.

- In Stata, use the command

```
xi: regress y x i.firm
```

Demeaning

- Another approach is to demean the variables across time.
- Consider

$$\begin{aligned} y_{it} - \bar{y}_i &= \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_{it}) \\ \tilde{y}_{it} &= \beta\tilde{x}_{it} + \tilde{u}_{it} \end{aligned}$$

  where notice that the mean variables (with bars over the top) are firm-dependent (note $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$).

- Why does this work? Because α_i is time invariant — its mean is the same as itself. Drops-out in the differencing.
- In Stata, use the command

```
xtreg y x, fe
```

  should offer the same results as the dummy approach.

Roadmap

Treatment Effects

- Difference in differences is a method used to estimate the impact of a treatment effect.
- Recall that simply taking a difference in the averages for treated and untreated samples leads to a selection bias term.
- The key is to take another difference!

Treatment Effects

- Consider a two-period example.
- In 1987 Arizona passed anti-takerover legislation, while Connecticut had not.
- Assume we have data pre and post reform: start of 1986 and end of 1987.
- Arizona firms are the treatment group and Connecticut firms are the control group.

Treatment Effects

- The regression model for the DD estimator is given by

$$y = \beta_0 + \beta_1 d \times p + \beta_2 d + \beta_3 p + u$$

  where

  - d is the treatment assignment variable equal to 1 if an Arizona firm and 0 if a Connecticut firm.
  - p is the post-treatment indicator equal to 1 if datum is from 1987 (post-reform) and 0 if from 1986 (pre-reform).

- β_2 captures differences across the two states, while β_3 captures differences across time.

- β_1 is our object of interest here: how did the policy change affect the Arizona firms?

Treatment Effects

- What are the possible combinations of outcomes?
  - $(d = 0) \wedge (p = 0) \Rightarrow y = \beta_0 + u.$
  - $(d = 1) \wedge (p = 0) \Rightarrow y = \beta_0 + \beta_2 + u.$
  - $(d = 0) \wedge (p = 1) \Rightarrow y = \beta_0 + \beta_3 + u.$
  - $(d = 1) \wedge (p = 1) \Rightarrow y = \beta_0 + \beta_1 + \beta_2 + \beta_3 + u.$

where $\wedge$ is shorthand for "and".

Treatment Effects

- Assume that $\mathbb{E}[u|p,d] = 0$: that is, the expectation of the residual is zero, irrespective of the values of d and p.

- Conditional expectations are all given then by
  - $\mathbb{E}[y|d=0,p=0] = \beta_0$
  - $\mathbb{E}[y|d=1,p=0] = \beta_0 + \beta_2$
  - $\mathbb{E}[y|d=0,p=1] = \beta_0 + \beta_3$
  - $\mathbb{E}[y|d=1,p=1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$

- So our DD estimator is given by

$$(\mathbb{E}[y|d=1,p=1] - \mathbb{E}[y|d=0,p=1])$$
$$-(\mathbb{E}[y|d=1,p=0] - \mathbb{E}[y|d=0,p=0]) = (\beta_1+\beta_2) - (\beta_2) = \beta_1$$

Treatment Effects

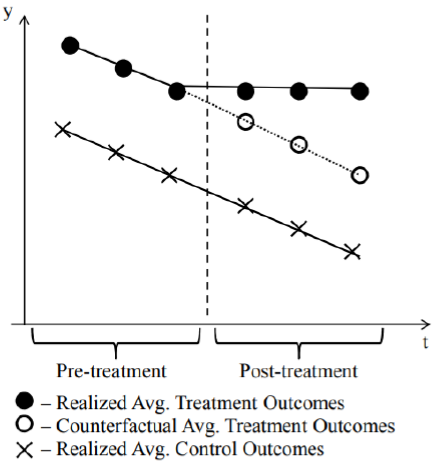
- Approximate this difference using sample averages!

Identifying assumption

- For this to work, we need the parallel trends assumption to hold.
- Means that the time-trend the two groups would have followed would be the same in the absence of the treatment.
- In the counter-factual, the trend would have been the same for the treatment and control groups.
- Then we can attribute the difference to the treatment effect.

Identifying assumption

Figure 1: **Difference-in-Differences Intuition**



Identifying assumption: robustness tests

- We don't observe the counterfactual outcome for the treatment group.
- Test: repeat the DD estimation for previous years (where the treatment was not present).
- Estimated treatment effect should be no different from zero statistically.
- Other tests possible as well.

Roadmap

Summary

- Linear regressions are easy to implement in Stata and can be very informative.
- Selection bias and endogeneity bias can be issues though.
- To deal with endogeneity: IV, panel data or DiD.