Corporate Finance

Lecture 2a: Review of Econometrics

Tong Li (李通)

School of Economics, Xiamen University
Wang Yanan Institute for Studies in Economics (WISE), Xiamen University

Email: litong@xmu.edu.cn

Outline for This Lecture

1. Econometrics and Corporate Finance
2. OLS Regressions
3. Correlation vs. Causality
4. Endogeneity and Solutions

What is Econometrics?

- Application of statistical methods to economic data
- Essential for analyzing financial data and testing theories
- Bridges the gap between theory and real-world data

Types of Economic Data in Econometric Analysis

1. Cross-Sectional Data
   - Data collected at a single point in time from multiple subjects (e.g., firms, individuals, countries)
   - Example: sales data from 100 firms in 2023

Types of Economic Data in Econometric Analysis

1. Cross-Sectional Data
    - Data collected at a single point in time from multiple subjects (e.g., firms, individuals, countries)
    - Example: sales data from 100 firms in 2023
2. Time Series Data
    - Data collected on a single subject over multiple time periods
    - Example: Annual sales of a firm from 2011 to 2020

Types of Economic Data in Econometric Analysis

1. Cross-Sectional Data
   - Data collected at a single point in time from multiple subjects (e.g., firms, individuals, countries)
   - Example: sales data from 100 firms in 2023

2. Time Series Data
   - Data collected on a single subject over multiple time periods
   - Example: Annual sales of a firm from 2011 to 2020

3. Panel Data
   - Data that combines both cross-sectional and time series elements
   - Tracks multiple subjects across time periods
   - Example: Annual sales of 100 firms over 10 years

Types of Economic Data in Econometric Analysis

1. Cross-Sectional Data
   - Data collected at a single point in time from multiple subjects (e.g., firms, individuals, countries)
   - Example: sales data from 100 firms in 2023
2. Time Series Data
   - Data collected on a single subject over multiple time periods
   - Example: Annual sales of a firm from 2011 to 2020
3. Panel Data
   - Data that combines both cross-sectional and time series elements
   - Tracks multiple subjects across time periods
   - Example: Annual sales of 100 firms over 10 years
4. Pooled Cross-Sectional Data
   - Cross-sectional data collected at different time points, but not necessarily following the same subjects
   - Example: Household income surveys from 2010 and 2020, but with different households

Importance of Econometrics in Corporate Finance

- Common Questions:
  - How does corporate capital structure affect firm value?
  - What is the impact of investment decisions on performance?
- Need for Empirical Evidence:
  - Validate theoretical models
  - Support decision-making

Ordinary Least Squares (OLS) Regression

- Basic Form of OLS:

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

  - y: Dependent variable
  - x: Independent variables
  - ϵ: Error term

Ordinary Least Squares (OLS) Regression

- Basic Form of OLS:

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

  - y: Dependent variable
  - x: Independent variables
  - ϵ: Error term

- Key assumptions of OLS:
  - **Linearity**: Relationship between x and y is linear
  - **Exogeneity**: x is uncorrelated with the error term ϵ
  - **No Omitted Variable Bias**
  - **No Perfect Multicollinearity**: Independent variables are not perfectly correlated
  - **Homoscedasticity**: Constant variance of errors

OLS in Corporate Finance: One Example (Gulen and Ion, RFS 2016)

- Background
  - Businesses often face a significant amount of uncertainty regarding the timing, content, and potential impact of policy decisions.
- Research Question:
  - Does policy-related uncertainty affect corporate investment decisions?

OLS in Corporate Finance: One Example (Gulen and Ion, RFS 2016)

- Background
  - Businesses often face a significant amount of uncertainty regarding the timing, content, and potential impact of policy decisions.
- Research Question:
  - Does policy-related uncertainty affect corporate investment decisions?
- OLS Regression Equation:

$$\frac{CAPX_{i,t+n}}{TA_{i,t+n-1}} = \alpha_i + \beta_1 PU_{i,t} + \beta_2 TQ_{i,t} + \beta_3 \frac{CF_{i,t}}{TA_{i,t-1}} + \beta_4 SG_{i,t} + \delta M_t + QRT_t + \epsilon_{i,t+n}$$

  - $\frac{CAPX_{i,t+n}}{TA_{i,t+n-1}}$: capital expenditure (scaled by total assets) of firm i in year $t+n$;
  - $PU_{i,t}$: policy uncertainty faced by firm i in year t;
  - $TQ_{i,t}$, $\frac{CF_{i,t}}{TA_{i,t-1}}$, and $SG_{i,t}$: firm-level control variables;
  - M_t: macroeconomic factors, e.g., GDP growth;
  - α_i: firm fixed effects; QRT_t: quarter fixed effects

OLS in Corporate Finance: One Example (Gulen and Ion, RFS 2016)

Table 2
Policy uncertainty and capital investment

Dependent variable: CAPX/Total assets	Panel A : Overall policy uncertainty index			
	(1)	(2)	(3)	(4)
Policy uncertainty	−0.168***	−0.171***	−0.179***	−0.174***
	(−6.67)	(−6.75)	(−7.04)	(−7.11)
Tobin's q	0.169***	0.159***	0.143***	0.128***
	(24.43)	(23.63)	(22.18)	(18.65)
Cash flow	0.0258***	0.0359***	0.0379***	0.0347***
	(10.04)	(13.70)	(14.22)	(13.31)
Sales growth	0.0409***	0.0448***	0.0397***	0.0305***
	(14.69)	(15.52)	(13.84)	(10.54)
GDP growth	0.0111	0.0199**	0.0225**	0.0324***
	(1.43)	(2.31)	(2.48)	(3.39)
Election indicator	0.00448	−0.00517	−0.0155	−0.0251
	(0.28)	(−0.33)	(−0.90)	(−1.23)
N	424,785	412,621	401,744	392,679
R-squared	0.039	0.038	0.033	0.028
Firm fixed effects	yes	yes	yes	yes
Quarter dummies	yes	yes	yes	yes
Cluster by firm	yes	yes	yes	yes
Cluster by quarter	yes	yes	yes	yes

Statistical Softwares for Econometric Analysis

- **Stata**
- R
- Python
- SAS
- ...

Correlation vs. Causality

- As researchers, we are interested in making **causal** statements
  - What is the effect of a change in corporate taxes on firms' leverage choices?
  - Does hiring a female CEO lead to better environmental performance of the firm?
- Avoid using "associated" or "correlated" to describe the relation between two variables
- Casual statements are essential for decision making (e.g., policymakers, corporate managers).

What do we mean by causality?

- If our linear model is the following

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

  and we want to infer β_1 as the causal effect of x_1 on y, holding all else equal, then we need to make the following assumptions:

  1. $E(\epsilon) = 0$
  2. $E(\epsilon | x_1, \cdots, x_k) = E(\epsilon)$

- This is "conditional mean independence" (CIM).
- Generally speaking, you need the estimation error to be uncorrelated with all the x.
- CIM is violated whenever an independent variable (x) is correlated with the error term (ϵ).

What is Endogeneity?

- Endogeneity is correlation between the error term (ϵ) and explanatory variables (x) in a regression

$$y = \alpha + \beta x + \epsilon$$

What Causes Endogeneity?

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

$$wage = \alpha + \beta education + \epsilon$$

- Hypothesis: $\beta > 0$, More education leads to higher wages.

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

$$wage = \alpha + \beta education + \epsilon$$

   - Hypothesis: $\beta > 0$, More education leads to higher wages.
   - BUT, people with higher ability (unobserved and omitted) tend to earn more and they likely get more education.

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

$$wage = \alpha + \beta education + \epsilon$$

   - Hypothesis: $\beta > 0$, More education leads to higher wages.
   - BUT, people with higher ability (unobserved and omitted) tend to earn more and they likely get more education.

2. **Measurement Error**: variables are misspecified or suffer from inaccurate data collection

$$Investment = \alpha + \beta_1 q + \beta_2 CashFlow + \epsilon$$

   - Hypothesis: $\beta_2 > 0$, more internal funds mitigate financial constraints and lead firms to invest more.

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

$$wage = \alpha + \beta education + \epsilon$$

   - Hypothesis: $\beta > 0$, More education leads to higher wages.
   - BUT, people with higher ability (unobserved and omitted) tend to earn more and they likely get more education.

2. **Measurement Error**: variables are misspecified or suffer from inaccurate data collection

$$Investment = \alpha + \beta_1 q + \beta_2 CashFlow + \epsilon$$

   - Hypothesis: $\beta_2 > 0$, more internal funds mitigate financial constraints and lead firms to invest more.
   - BUT, firms with more cash have better investment opportunities (mismeasured by Tobin's q) and they investment more.

What Causes Endogeneity?

1. **Omitted Variables**: missing factors affecting y

$$wage = \alpha + \beta education + \epsilon$$

   - Hypothesis: $\beta > 0$, More education leads to higher wages.
   - BUT, people with higher ability (unobserved and omitted) tend to earn more and they likely get more education.

2. **Measurement Error**: variables are misspecified or suffer from inaccurate data collection

$$Investment = \alpha + \beta_1 q + \beta_2 CashFlow + \epsilon$$

   - Hypothesis: $\beta_2 > 0$, more internal funds mitigate financial constraints and lead firms to invest more.
   - BUT, firms with more cash have better investment opportunities (mismeasured by Tobin's q) and they investment more.

3. **Reverse Causality (Simultaneity)**: y may also affect x

$$Yield = \alpha + \beta Amount + \epsilon$$

$$Amount = \theta + \delta Yield + \eta$$

Consequences of Endogeneity

- Statistically speaking, regression parameters cannot be identified.
  - OLS estimates will not be estimates of what we want.
- Practically speaking, it is very hard (impossible) to interpret the results.
  - We cannot rule out alternative explanations for findings.

Common Tools to Establish Causality

1. Instrumental Variables
2. Difference-in-Differences (natural/qusi-natural experiments)
3. Regression Discontinuity Design

Further Reading

- Stock & Watson, Introduction to Econometrics
- Angrist & Pischke, Mostly Harmless Econometrics: An Empiricist's Companion
- Rodríguez, Stata Tutorial

Reminder

- Please upload Learning Objectives by **next Sunday** (Sep 15th)
- No lecture next Sunday (Sep 15th)