Class 17 Case Study: Estimating Causal Network Effects for Platform Businesses Using Instrumental Variables

Dr. Wei Miao

UCL School of Management

November 26, 2025

Section 1

Causal Questions for Platform Businesses

Class objectives:

- Understand the importance of causal inference for platform businesses.
- Learn how to estimate causal effects using instrumental variables with an application to food delivery platforms.

Causal Questions for Platform Businesses

- Platform businesses often need to answer critical causal questions to optimise their operations:
  - **Measuring network effects**: How does increasing supply (restaurants) affect demand (orders)?
  - **Pricing**: How does surge pricing affect consumer demand and driver supply?
- When relying on secondary non-experimental data, these questions often face **endogeneity** challenges that require careful empirical strategies.

Section 2

Case Study

Case Study

Core case question: How to estimate the cross network effect (CNE) of restaurant density on total orders using historical data?[1]

- We can run a linear regression model on UberEats' historical data, where the dependent variable is the total number of orders; the key explanatory variable is the number of restaurants in a postcode.

$$\text{TotalOrders} = \beta_0 + \beta_1 \text{NumberRestaurants} + \varepsilon$$

- Is there endogeneity in this model?[2]

[1] Direct network effect (DNE) is defined as the network effect that arises from the number of users on the same side of the platform. For instance, more users on UberEats will lead to more new users joining the platform.

[2] Answers are available on the HTML version.

Postcode Level Data

- Aggregated transaction data from London postcodes (e.g., E1, W1).
  - **total_orders**: Outcome variable (Y)
  - **n_restaurants**: Endogenous explanatory variable (X)
  - **avg_income**: Control variable - Average income in the postcode

	postcode_id	postcode_area	year	week	n_restaurants	hygiene_inspections	avg_income	total_orders
1	1	N13	2024	1	52	5	25.43575	2105
2	2	WC12	2024	1	42	16	65.71556	2009
3	3	WC16	2024	1	21	13	36.47918	931
4	4	W7	2024	1	45	11	40.59615	1980
5	5	NW2	2024	1	21	19	30.44267	1519

Section 3

Empirical Strategy

OLS Linear Regression

$$TotalOrders_i = \beta_0 + \beta_1 NumberRestaurants_i + \beta_2 AvgIncome_i + \varepsilon_i$$

- **Exercise**: Run the linear regression model controlling for income. Discuss the direction of the bias.

Instrumental Variables Regression

Can you come up with a valid instrument for the number of restaurants?

- Instruments that satisfy (1) relevance, (2) exogeneity, and (3) exclusion restriction

Answer

- **Instrument (Z)**: Number of Hygiene Inspections
- **Logic**:
  - Random inspections increase costs/risk for restaurants, reducing supply (Relevance).
  - Random inspections are conducted by local council and are beyond the control of restaurants (Exogeneity).
  - Inspections are bureaucratic processes unrelated to consumer hunger and thus should not directly affect total orders (Exclusion Restriction).

Two-Stage Least Squares (2SLS) Estimation: First Stage

- First stage: Regress endogenous variable on the instrument **and controls**.

$$NumberRestaurants_i = \pi_0 + \pi_1 HygieneInspections_i + \pi_2 AvgIncome_i + \varepsilon_i$$

- Check for instrument relevance: Is π_1 significantly different from zero? (Expect negative sign).

Two-Stage Least Squares (2SLS) Estimation: Second Stage

- Second stage: Regress outcome on predicted restaurant count **and controls**.

$$TotalOrders_i = \beta_0 + \beta_1 Number\widehat{Restaurants}_i + \beta_2 AvgIncome_i + \varepsilon_i$$

- The coefficient β_1 is the causal effect of restaurant supply on orders.
- **Exercise**: Run the 2SLS regression model. Compare the coefficient with OLS.

TSLS in One Step

- We can run the 2SLS regression in one step using the `feols` function in the `fixest` package.

```
model_iv_2nd <- feols(
    fml = total_orders ~ avg_income |
        n_restaurants ~ hygiene_inspections,
    data = data_london
)
```

- The `fml` syntax is `fml = Y ~ Control | X ~ Z`, where `Y` is the outcome variable, `Control` is the control variable, `X` is the endogenous variable, and `Z` is the instrument.

The Cost of Getting It Wrong (Back-of-the-Envelope)

- Suppose we want to know if we should spend £2,000 to onboard 10 new restaurants in a postcode. Margin per order is £5.
- **Biased OLS Estimate**: $\hat{\beta}_{OLS} = 44.2$ orders/restaurant.
  - Projected Gain: $10 \times 44.2 \times £5 = £2,210$.
  - Profit: $£2,210 - £2,000 = £\mathbf{210}$.
  - **Decision**: Invest.
- **True Causal Effect (IV)**: $\hat{\beta}_{IV} = 13.2$ orders/restaurant.
  - Projected Gain: $10 \times 13.2 \times £5 = £660$.
  - Profit: $£660 - £2,000 = £ - \mathbf{1,340}$.
  - **Decision**: Do Not Invest.
- **Consequence**: Relying on OLS leads to a **bad investment decision**. We would invest in a campaign that actually loses money (or miss a profitable one, depending on bias direction).

Section 4

Instrumental Variable in A/B Testing (Optional)

A/B Testing with an Encouragement Design

- In experimental settings (A/B testing), IVs are used to handle **imperfect compliance**. This is often called an **Encouragement Design**.
- **Intended Treatment (Z)**: Random assignment to a treatment through A/B testing (e.g., a subsidy offer to restaurants).
- **Outcome (Y)**: Business metric (e.g., orders).

$$Y = \beta_0 + \beta_1 Z + \varepsilon$$

- β_1 is the **intent-to-treat effect** of the treatment on the outcome.

LATE: Local Average Treatment Effect

- **Actual Treatment (X)**: Actual uptake of the treatment (e.g., joining the platform). However, the following regression is endogenous because those who receive the treatment and join the platform are fundamentally different from those who receive the treatment but do not join the platform.

$$Y = \gamma_0 + \gamma_1 X + \varepsilon$$

- Since we cannot force participation, we use the random assignment (Z) as an instrument for actual participation (X).
- The IV estimator recovers the **Local Average Treatment Effect (LATE)**: the effect on those induced to participate by the encouragement.

Example 1: Sign Up Bonus for New Customers

- **Context**: UberEats offers a random £10 sign-up bonus to a subset of potential customers to encourage them to download the app.
- **Causal Question**: What is the effect of **App Download (X)** on **Customer Lifetime Value (Y)**?
- **Endogeneity**: Customers who download the app organically are likely more interested in food delivery (higher CLV) than those who don't.
- **Instrumental Variable**:
  - **Instrument (Z)**: Random Assignment to receive the bonus offer (Intent-to-Treat).
  - **Treatment (X)**: Did the customer download the app?
  - **Outcome (Y)**: CLV (orders placed in the next 6 months).
- **IV Logic**: The bonus is randomly assigned (exogenous). It increases the probability of downloading the app (relevance). It affects CLV only through the app usage (exclusion).
- **LATE**: The effect of the app on CLV for those customers who *only* downloaded it because of the £10 bonus (the marginal customers).

Example 2: Apple's App Tracking Transparency (After-Class Reading)

- **Context**: Apple introduced ATT, requiring apps to ask users for permission to track them across other apps and websites.
- **Causal Question**: What is the effect of **Tracking Permission (X)** on **Ad Effectiveness (Y)**?
- **Endogeneity**: Users who opt-in to tracking are fundamentally different from those who opt-out (selection bias). We cannot simply compare opt-in vs. opt-out users.
- **Instrumental Variable Strategy**:
  - **Instrument (Z)**: The **Policy Change** itself (Pre-ATT vs. Post-ATT, or random rollout). The policy "encourages" (or forces) users to make a choice, changing the probability of being tracked.
- **IV Logic**: The policy change is exogenous (determined by Apple). It strongly affects tracking rates (Relevance). It affects ad effectiveness primarily through the loss of tracking data (Exclusion).
- **LATE**: The IV estimate gives us the effect of tracking on ad effectiveness for the "compliers" (users whose tracking status changed due to the policy).

After-Class Reading

- (highly recommended) Encouragement Designs and Instrumental Variables for A/B Testing at Spotify
- Chu, Junhong, and Puneet Manchanda. 2016. "Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms." Marketing Science 35 (6): 870–93.