

# Session 9

## Session 9-1 도구변수

- 도구변수는 연구디자인이 어려울 때의 대안
- 

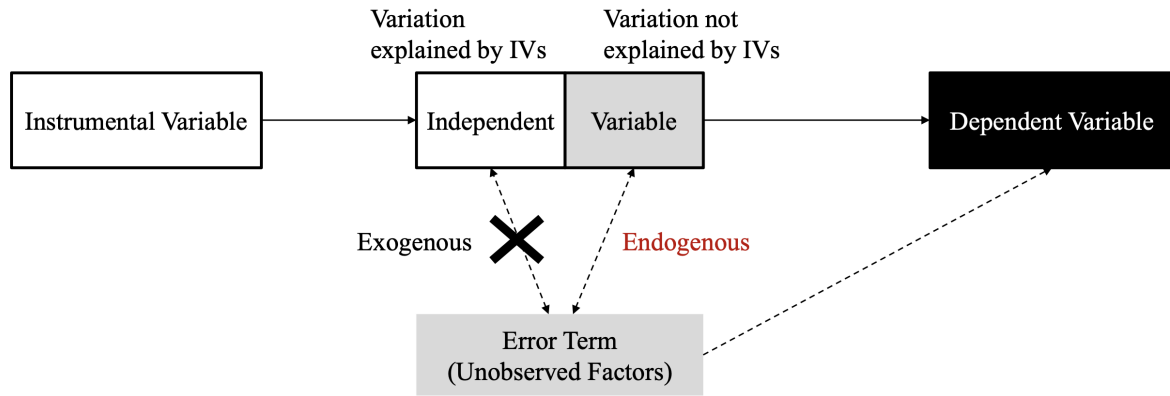
### Three Perspectives on Causation

- Potential Outcomes Framework : 인과추론을 방해하는 요인은 selection bias. 이는 treatment와 control 그룹이 서로 비교가능하지 않아서 생기는 차이.
- Structural Causal Model : 인과추론을 방해하는 요인은 backdoor path. do(X)를 통해 해결함.
- Statistics(: Regression) : 인과추론은 통계적 비편향성으로 정의하게 되고 이를 방해하는 것은 내생성(Endogeneity). independent 변수가 error와 correlation이 있는 것이 내생성.

### Endogeneity in Regression

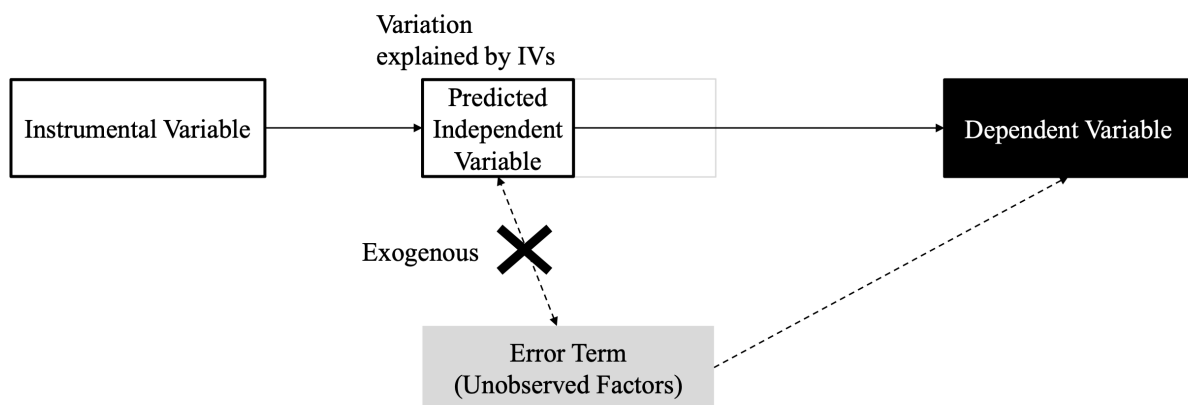
- Exogenous : 원인변수와 에러와 상관관계가 없는 것.
- regression 하에서 추정 결과를 인과관계로 해석하기 위해서는 원인변수와 에러와 상관관계가 없어야 한다.
- 에러는 우리가 관찰할 수 없는 모든 요인이 포함되는 개념.
- 현실적으로 모든것은 endogenous. 즉, 단순회귀로는 이 문제를 해결할 수 없다.
- 이를 위한 도구가 도구변수.

### Taking the Selection Bias Out: Instrumental Variable(IV)



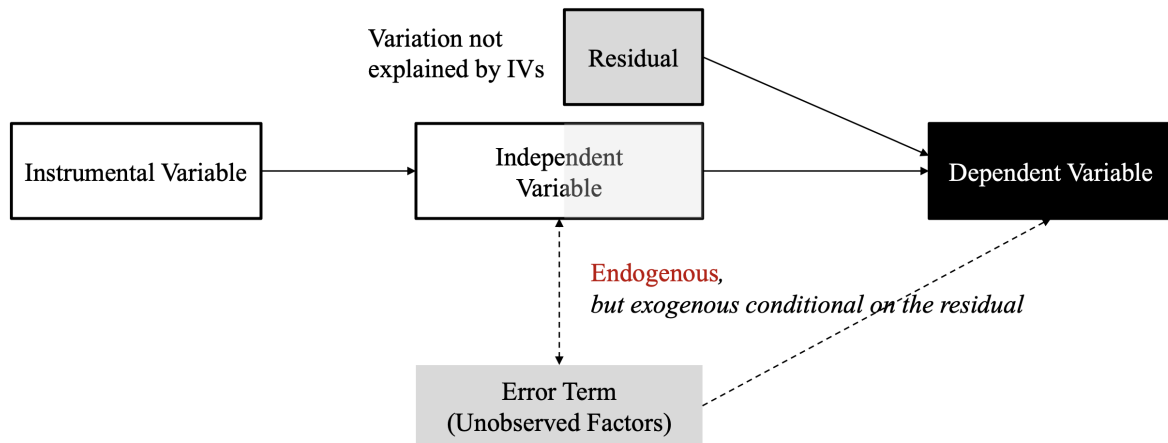
- 원인변수 중에 error term과 상관관계를 갖는 부분을 인위적으로 도려내자는 것이 도구 변수의 기본 아이디어.

## First Approach: Two-Stage Least Squares



1. 도구변수를 사용해서 원인변수를 예측한다. ( $X = a_0 + a_1 Z + \epsilon'$ )
2. 도구변수가 외생성 부분을 잘 예측한다는 가정하에서 그 부분만 따로 떼서 분석하면 비편향 추정이 되고 인과관계를 해석할 수 있다. ( $Y = b_0 + b_1 X + \epsilon \rightarrow Y = b_0 + b_1 \hat{X} + \epsilon''$ )

## Second Approach: Control Function



- 원인변수를 그대로 사용하되 인과추론에 문제가 되는 내생성 부분만 control하자.
- $X = a_0 + a_1 Z + \epsilon'$ 에서 residual을 control그룹으로 추가한다.

- Basic idea of control function

$$\begin{aligned} X &= \gamma Z + \underbrace{(v)}_{\text{내생성}} \\ Y &= \beta X + u \end{aligned} \quad \left. \vphantom{\begin{aligned} X &= \gamma Z + (v) \\ Y &= \beta X + u \end{aligned}} \right\} \text{Error terms are correlated.}$$

$X$ 는  $Z$ 과  $v$ 가 독립이면  
 $\gamma$ 는  $\gamma$ 가 된다

Control Function

$$E(Y|Z, v) = \beta(X|Z, v) + E(u|v) = \beta X + \rho v$$

$$Y = \beta X + \rho v + \varepsilon$$

$\rho$  is the correlation between  $u$  and  $v$ .

원인은  $E(u|Z, v)$   
 도구변수는 이리됨과 상관 관계가 있음.

## Identification Assumptions for IV

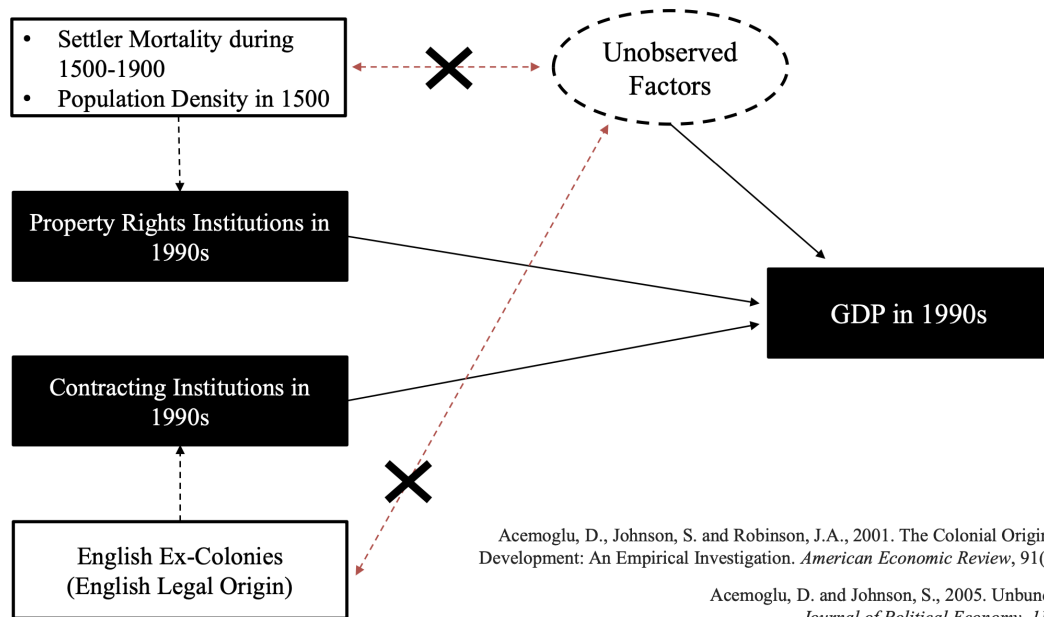
- 도구변수가 예측하는 부분이 외생성 이어야한다. 즉, 에러터ム과 상관관계가 없어야한다.
- 도구변수는 원인변수에 대해 충분히 설명할 수 있어야한다.

## Local Average Treatment Effect (LATE)

- 부분적 인과관계는 도구변수의 한계점으로 볼 수 있다.
- 이러한 부분적 인과관계의 효과를 LATE라고 한다.
- 도구변수에 의해서만 treatment가 변하는 것을 Compliers라고 한다.
- LATE는 monotonicity가정이 필요하다. 즉 도구변수가 없을 때 treatment가 변하는 Defiers그룹이 없어야 한다.

## Session 9-2 도구 변수의 활용 사례

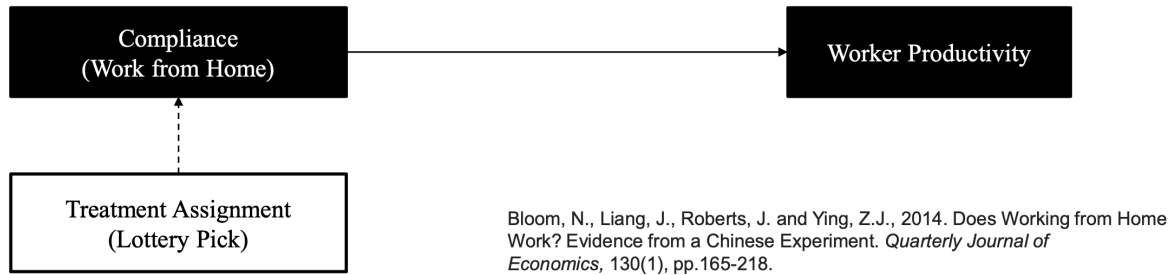
### Ex1. Institution and Economic Growth



- 전세계를 봤을 때 쌍둥이 도시를 찾기 어렵기 때문에 준실험방법을 하기 어렵고 도구변수가 그 대안이 될 수 있다.
- 위 그림은 사유재산제도와 계약제도가 경제성장에 미치는 영향에 대한 그림.
  - 경제성장에 미치는 요인은 굉장히 많고 관찰할 수 없다. 즉, 위 원인변수들은 내생적이다.
  - 사유재산제도에 대한 도구변수로는 제국주의시절 식민지의 사망률, 인구밀도를 활용, 계약제도에는 식민 지배여부를 도구변수로 사용한다.

## Ex2. Noncompliance in Randomized Experiments

- assign된 treatment그룹과 control 그룹의 차이를 비교하므로서 추정하고자 의도한 효과를 Intention to treat(ITT)라고 한다.
- 하지만 경우에 따라 treatment를 받았다고 해서 참가자들이 준수하지 않을 수도 있다. 이를 Noncompliance문제라고 한다.
- Noncompliance 문제를 해결하기 위한 방법은 treatment assignment를 도구변수로 활용하는 것.



재택근무가 근로자의 생산성에 미치는 영향에 대한 실험

- 랜덤으로 뽑기를 통해 재택 여부를 결정. 여기서 뽑힌 사람과 안뽑힌 사람을 비교하는 것이 ITT.
- 하지만 재택이 가능하다해도 재택을 안할수도 있다. 따라서 random assignment만으로 비교하는 것은 무리가 있다. 이를 Noncompliance.
- 이럴경우 treatment assignment를 도구변수로 두면 실제 재택근무만의 효과만을 분석할 수 있다.

## Session 9-3 도구 변수의 활용 팁

### How to Report IV Analysis

- step1. 도구변수에 대해 통계적 테스트를 진행
- step2. 도구변수에 대한 이론적 타당성을 정당화
- step3. 도구변수에 대한 민감도 테스트를 진행
- step4. LATE인지 ATE인지 명확한 분석

### 통계적 테스트는 필요조건이지만 충분조건은 아니다

- relevance에 대한 가정은 통계적으로 검증할 수 있다.
- 하지만, exclusion restriction, exogeneity assumption은 검증할 수 없다. 말그대로 가정이기 때문
  - 존재하는 test들은 굉장히 불안정함. 도구변수의 적절성을 검증하기 위해 도구변수의 적절성을 가정하기 때문
    - Hausman test : endogeneity test, H0: OLS와 IV의 추정치의 차이가 유의하지 않다. 이는 도구변수가 완전하다고 가정하에서만 성립하는 test.
    - Sargan-Hansen J test : exclusion restriction, exogeneity assumption에 대한 test. H0: k-1개의 도구변수를 사용하는 오차항과 다른 도구변수의 상관

관계가 없다.