Stronger Exogeneity Assumptions
Lecture for Economics 240A

Douglas G. Steigerwald

UC Santa Barbara

September 2011
Do certain mutual fund managers produce consistently high returns?

- **excess returns**: fund return - average of all fund returns
- Morningstar calculates annual excess return for 10,000 funds
- How would you determine if certain fund managers produce consistently high returns?
Stochastic Model
Notation for Indicator Variables

- $n$ funds indexed by $i$
  - $R_i$ - fund return
  - $E_i$ - fund excess return \[ E_i \equiv R_i - \frac{1}{n} \sum_{j=1}^{n} R_j \]
  - $H_i$ - indicator for positive excess return in previous year
    \[ H_i = \begin{cases} 
    1 & \text{if } E_i > 0 \text{ in previous year} \\
    0 & \text{if } E_i \leq 0 \text{ in previous year} 
    \end{cases} \]
  - compact notation \[ H_i = 1 \{ E_i > 0 \text{ in previous year} \} \]
  - $U_i$ - other forces that affect excess returns

Stochastic Model

\[ E_i = \beta_0 + \beta_1 H_i + U_i \]

- $\beta_0$ - expected excess return for a fund
- $\beta_1$ - change in expected excess return for a fund with a positive excess return the previous year

Previous high returns indicate future high returns: $\beta_1 > 0$
Exogeneity Assumptions

\[ X_i^T = (1 \ H_i) \]

1. Standard identification assumption

\[ \mathbb{E} (X_i U_i) = 0 \]

2. Stronger identification assumption

\[ \mathbb{E} (U_i | X) = 0 \]

3. Strongest identification assumption

\[ U_i \text{ is statistically independent of } X \text{ and } \mathbb{E} (U_i) = 0 \]

\[ 3 \Rightarrow 2 \Rightarrow 1 \]
Framework: Conditional Expectation

- Suppose $U$ takes $J$ values

\[
\mathbb{E}(U|X = x) = \sum_{j=1}^{J} u_j \cdot \mathbb{P}(U = u_j|X = x)
\]

- the conditioning argument is $X = x$
- the expected value is of $U$ given that $X$ takes the value $x$
- the value $x$ affects the probability that $U$ takes each value

- Example: $X$ takes the value 1 for $H$ from a fair coin flip
  - if $X = 1$, then $U$ is the value from a flip of a second fair coin
  - if $X = 0$, then $U$ is the value from a flip of a coin for which $\mathbb{P}(H) = .25$

- What is $\mathbb{P}(U = 1|X = 1)$?
  - $1/2$

- What is $\mathbb{P}(U = 1|X = 0)$?
  - $1/4$
Linking Identification Assumptions

- Suppose $U$ takes $J$ values and $X$ takes $K$ values
- Fact
  \[ \mathbb{P}(U = u_j, X = x_k) = \mathbb{P}(U = u_j|X = x_k) \cdot \mathbb{P}(X = x_k) \]

\[
\mathbb{E}(XU) = \sum_{k=1}^{K} \left[ x_k \sum_{j=1}^{J} u_j \mathbb{P}(U = u_j, X = x_k) \right]
= \sum_{k=1}^{K} \left[ x_k \sum_{j=1}^{J} u_j \mathbb{P}(U = u_j|X = x_k) \mathbb{P}(X = x_k) \right]
= \sum_{k=1}^{K} \left[ x_k \mathbb{P}(X = x_k) \mathbb{E}(U|X = x_k) \right]
\]

- $\mathbb{E}(U|X) = 0 \Rightarrow \mathbb{E}(U|X = x_k) = 0$ for all values $x_k \Rightarrow \mathbb{E}(XU) = 0$
- $2 \Rightarrow 1$
Two events are statistically independent if the occurrence of one event provides no information about the occurrence of the other event.

- Events $A$ and $B$ independent implies
  \[ P(A|B) = P(A) \quad P(A \text{ and } B) = P(A)P(B) \]

- If two random variables ($X$ and $U$) are independent, then information about one ($X$) provides no information about the other ($U$)

- $X$ independent of $U$ implies
  \[ \mathbb{E}(U|X) = \mathbb{E}(U) \]

- 3 $\Rightarrow$ 2
Identification Assumption Implication
Correct Specification

\[ E_i = \beta_0 + \beta_1 H_i + U_i \quad (\ast) \]

- strong exogeneity \( \Rightarrow \beta_0 + \beta_1 H_i = \mathbb{E} (E_i | X) \)
  - correct model of conditional mean
  - needed to predict \( E_i \)

- weak exogeneity \( \Rightarrow \text{Cor} (H_i, U_i) = 0 \)
  - adequate to capture marginal effect of \( H_i \)

When do they differ?

\[ E_i = \beta_0 + \beta_1 H_i + \beta_2 H_i^2 + \tilde{U}_i \quad \text{and} \quad H_i \text{ is symmetric about 0} \]

- estimated model is \( (\ast) \) with \( U_i = \beta_2 H_i^2 + \tilde{U}_i \)
  - \( \mathbb{E} (U_i | X) = \beta_2 H_i^2 \neq 0 \)
  - \( \mathbb{E} (H_i U_i) = \beta_2 H_i^3 + H_i \tilde{U}_i = 0 \)
samples of funds in many studies

\[ b_1 \approx 0 \]

- assume \( \mathbb{E}(X_i; U_i) = 0 \)
  - conclude \( H_i \) doesn’t influence \( E_i \)
- assume \( \mathbb{E}(U_i|X) = 0 \)
  - conclude \( H_i \) doesn’t influence \( E_i \)
  - conclude \( b_0 \) (the sample average) is the predictor for excess return for a fund
Question: Worker Effort

How should firms pay workers to elicit maximum effort?

- Model in vector form

\[ E_i = X_i^T \beta + U_i \]
\[ X_i = \begin{pmatrix} 1 \\ W_i \\ C_i \end{pmatrix} \]

- What does \( \mathbb{E} (X_i U_i) = 0 \) imply?
  - wage and co-worker wage are uncorrelated with other factors that determine effort, so any omitted regressors are uncorrelated with wage and co-worker wage

- What can we estimate?
  - marginal effects of wage and co-worker wage

- What does \( \mathbb{E} (U_i | X) = 0 \) imply?
  - no omitted variables

- What can we estimate?
  - marginal effects and predict the effort of a worker