Markets, Incentives and Ethical Management

2. Refresher on Game Theory

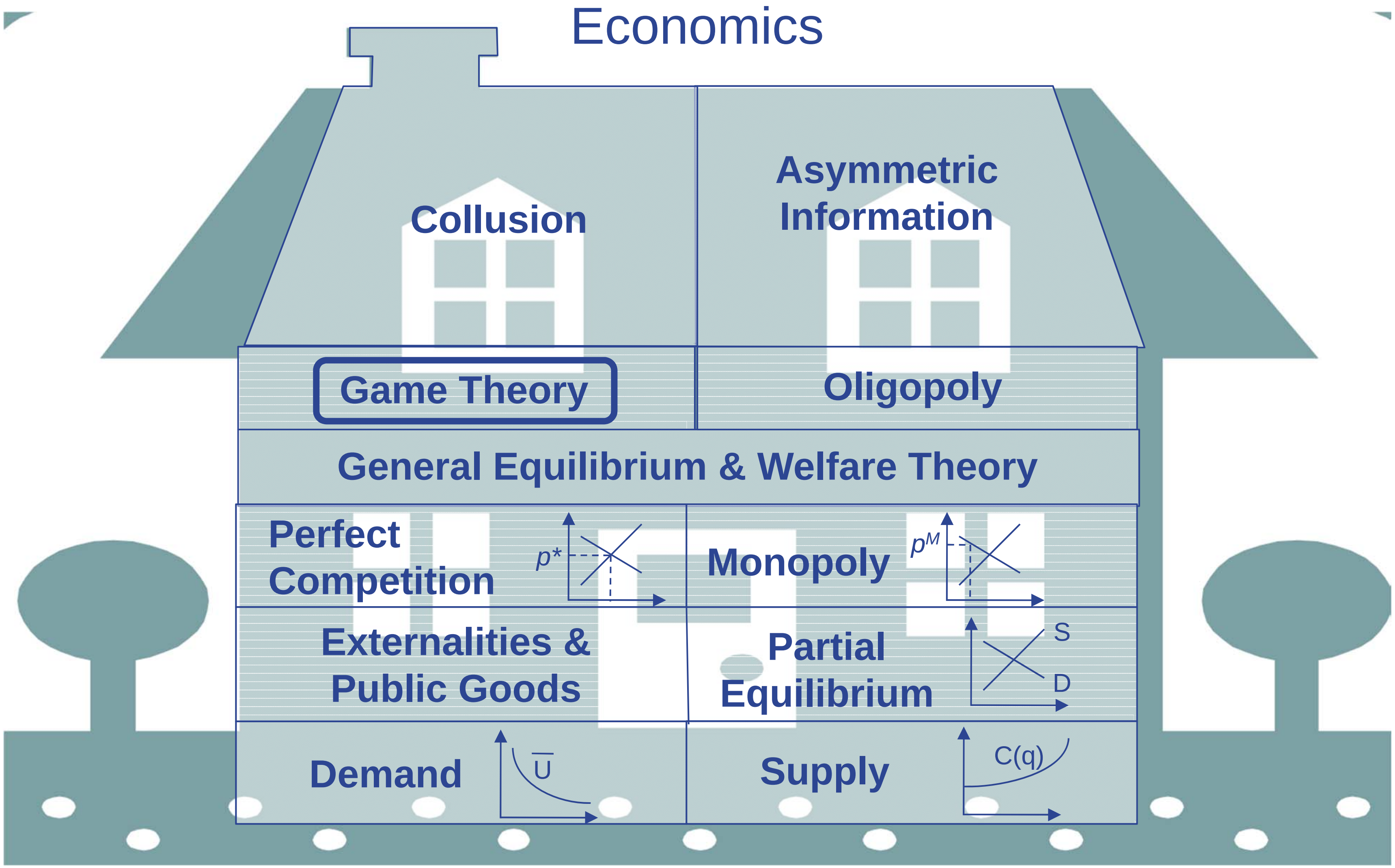
Prof. Dr. Markus Reisinger

Professor for Industrial Organization and Microeconomics

Frankfurt School of Finance & Management

m.reisinger@fs.de

Economics comes from Greek “oikos” which means “household”.



Aim of today's lecture

Aims

1. Refresh your knowledge on static games.
2. Refresh your knowledge on dynamic games.
3. See how game theory can be applied by using the example of an auction.

1	Static Games
2	Sequential Games
3	Application: An Auction

We try to find a “solution” for a game as a forecast for the outcome of the strategic interaction.

Example: Advertising Game

Situation

- There are two (symmetric) firms offering soft drinks (e.g., Pepsi, Coke). Production costs are (for simplicity) zero.
- Firms can invest in advertising (for simplicity, assume that they can choose only between the investment levels zero and €2 m.)
- However, the total demand for soft drinks is fixed at € 20 m., i.e., advertising can only steal business from the competitor
- If no firm invests, each firm earns €10 m. If one firm invests, and the other does not, its revenues will increase to €15 m., while the other firms revenues fall to €5 m. (and vice versa).
- If both firms invest, they still share the market and obtain a revenue of €10 m. each.
- Question: Would you invest in advertising? Will firms invest in advertising?

We need to “translate” the economic situation into a game to be able to use game theoretic analysis.

Situation	Game
• There are two (symmetric) firms offering soft drinks (e.g., Pepsi, Coke). Production costs are zero. • Firms can invest in advertising (for simplicity, assume that they can choose only between the investment levels zero and €2 mn.) • However, the total demand for soft drinks is fixed at € 20 mn., i.e. advertising can only steal business from the competitor • If no firm invests, each firm earns €10 mn. If one firm invests, and the other does not, its revenues will increase to €15 mn., while the other firms revenues fall to €5 mn. (and vice versa).	• Players: Two identical firms. • Actions: “advertise”, “don’t advertise” • Timing: advertising decisions are made simultaneously • Information: Players know the “situation” and the other firm’s profit functions • Outcomes: (don’t, don’t) -> (10,10) (do, don’t) -> (13,5) (don’t, do) -> (5,13) (do, do) -> (8,8)

A simple way to depict a (static) game is a payoff matrix (normal form representation).

The Advertising Game

		Player 2	
		Don't Advertise	Advertise
Player 1	Don't Advertise	(10,10)	(5,13)
	Advertise	(13,5)	(8,8)

Payoff of player 1

Payoff of player 2

The concept that determines the equilibrium outcome is the “Nash equilibrium”.

John Forbes Nash (1928 - 2015)
Nobel laureate in economics sciences 1994

The key equilibrium concept is “Nash equilibrium”: All players select best responses.

Nash Equilibrium (NE)

In a Nash equilibrium, each player selects an action, which is the best response to what the other players have chosen, that is, each player acts optimally, given the other players’ choices.

Advertising Game

		Player 2	
		Don’t Advertise	Advertise
Player 1	Don’t Advertise	(10,10)	(5,13)
	Advertise	(13,5)	$\boxed{(8,8)}$

The equilibrium outcome of the advertising game is a prisoner‘s dilemma

		Player 2	
		don’t advertise A	advertise B
Player 1	don’t advertise	(10,10)	(5,13)
	advertise	(13,5)	(8,8)

- Prisonner‘s dilemma: both players would be better off when they credibly commit not to invest into advertising (both would get 10 instead of 8)
- **But**, actions have to be made simultaneously and since the market cannot be expanded, advertising induces a business stealing effect

➔ Investing into advertising is a dominant strategy for both players!

To identify a Nash equilibrium, we have to analyze each outcome individually.

How do we find a Nash equilibrium?

- Nash equilibrium is an inherently static concept (we will later see, how it can be applied in dynamic games).
- In principle, we have to check for each outcome (i.e., for each combination of strategic choices) whether or not each player acts optimally: if a player has a profitable deviation, the outcome is not a Nash equilibrium.

Games with discrete payoffs

		Player 2: A	Player 2: B
Player 1	a	(1,10)	(1,13)
	b	$\boxed{(3,4)}$	(2,3)

Games with continuous payoffs

- With continuous payoffs, we can often find the Nash equilibrium by using the first-order conditions
- FOC describe the best response given what the other players do
- Example of an advertising game on the next slide

An advertising game with continuous payoffs.

Advertising Game

Situation

- There are two symmetric firms. Each firm can invest to advertise its product. We denote by a_1 and a_2 the advertising level of firm 1 and 2, respectively.
- Advertising costs $(a_i)^2$ for firm i =1,2.
- Benefit from advertising: increase in demand.
- However, if the rival advertises more, demand falls.
- Marginal gain from advertising is also lower if rival advertises more.
- Overall demand of firm 1 is given by:

$$1 + a_1 - \frac{1}{2}a_2 - a_1 a_2$$

(demand of firm 2 is the same with indices reversed)

- Price of the product equals 1.
- Question: What is the optimal advertising level of each firm?

Maximization of the profit function gives the "best response" function for each firm.

Solving the Advertising Game

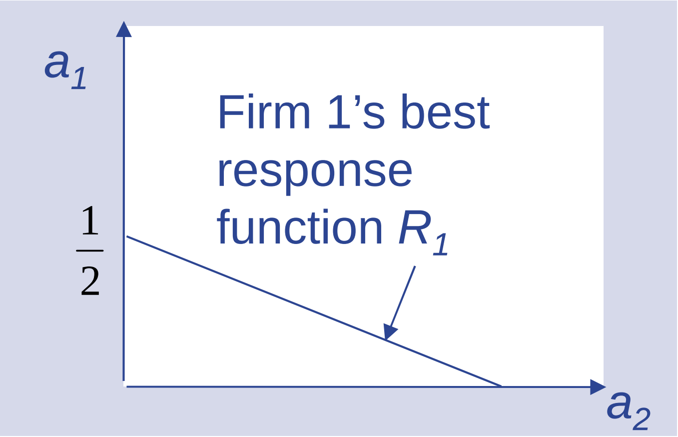
Firm 1 maximizes its profit function

(1) $$\pi_1 = 1 + a_1 - \frac{1}{2}a_2 - a_1 a_2 - a_1^2 \rightarrow \max_{a_1}!$$

We derive the first-order condition for (1):

$$\frac{\partial \pi_1}{\partial a_1} = 1 - a_2 - 2a_1 = 0 \qquad \Rightarrow \qquad a_1 = \frac{1 - a_2}{2} \qquad (2)$$

Equation (2) is firm 1's reaction (or "best response") function: it says: what is the optimal choice a_1 of firm 1 for any given level of a_2.



Reaction function is falling.

The equilibrium is given by the intersection of the best response functions.

Solving the Advertising Game (cont.)

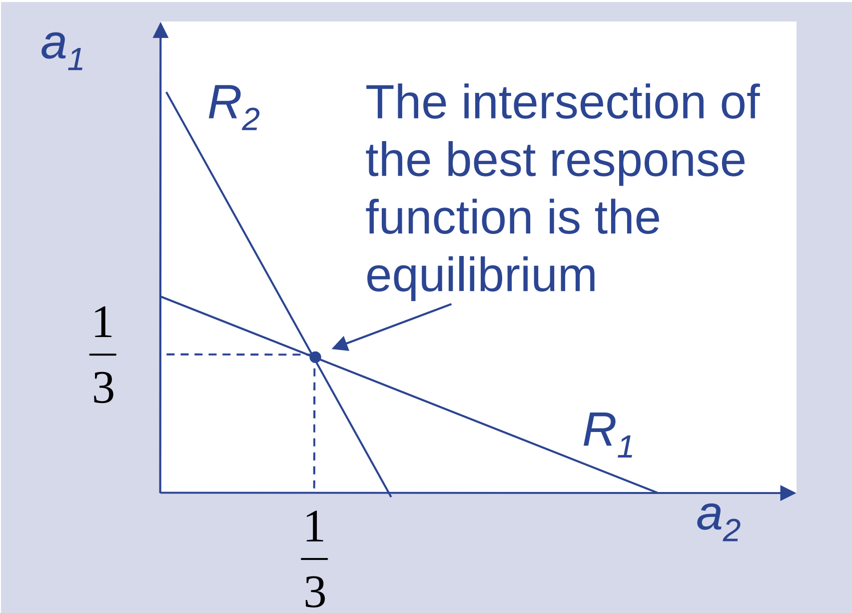
Following the same steps for firm 2: $\pi_2 = 1 + a_2 - \frac{1}{2}a_1 - a_2 a_1 - a_2^2 \to \max_{a_2}!$

Reaction function of firm 2: $a_2 = \frac{1 - a_1}{2}$ (3)

Equations (2) and (3) give the conditions for our equilibrium: if they are both satisfied, each firm plays a best response to the strategy of the other. Solving the system of two equations for the two unknowns, a_1 and a_2, yields:

$$a_1^* = a_2^* = \frac{1}{3}$$

This is the Nash-Equilibrium!



Equilibrium profits.

Solving the Advertising Game (cont.)

Equilibrium profit can be obtained by plugging the values of the Nash equilibrium advertising levels into the profit function:

$$\pi_1^* = 1 + \frac{1}{3} - \frac{1}{2} \cdot \frac{1}{3} - \frac{1}{3} \cdot \frac{1}{3} - \frac{1}{9} = \frac{17}{18}$$

As in the game with a discrete action set, the firms would be jointly better off by not investing in advertising. This would lead to a profit of 1.

1	Static Games
2	**Sequential Games**
3	Application: An Auction

We now want to allow for the case where the timing of strategic choices matters.

Example: Market Entry Game

Situation

- There are two pharmaceutical companies, which are about to develop a Corona vaccine and can serve the German market. One firm, denoted by firm A, is from Germany, the other firm, B, from England.
- Development and distribution requires huge investments of 10 each.
- The German firms has a first mover advantage: it can enter the market, “sink” the investment, and will then be able to supply the market.
- If B does not follow, A will get (gross of investment) profits of 12, leaving a total profit of 2.
- If B follows, each firms’ (gross) profits are 9 (and net profits are -1).
- If A does not enter but B enters, B gets (gross) profits of 12 if it enters.

Simultaneous Version of the Game.

The Market Entry Game

- Suppose that the game is played simultaneously.
- The game can then represented by the familiar strategic form.
- The game has two Nash equilibria.
- Are these two Nash equilibria also equilibria in the sequential game?
- If one firms enters, it is optimal for the other firm to stay out. But given that the other firm stays out, it is optimal for the first firm to enter.
- Therefore, the answer is yes!
- But is the Nash equilibrium in which firm B enters and firm A stays out a plausible one?

Strategic Form Representation

		B: enter	B: don't
A	enter	(-1,-1)	(2,0)
	don't	(0,2)	(0,0)

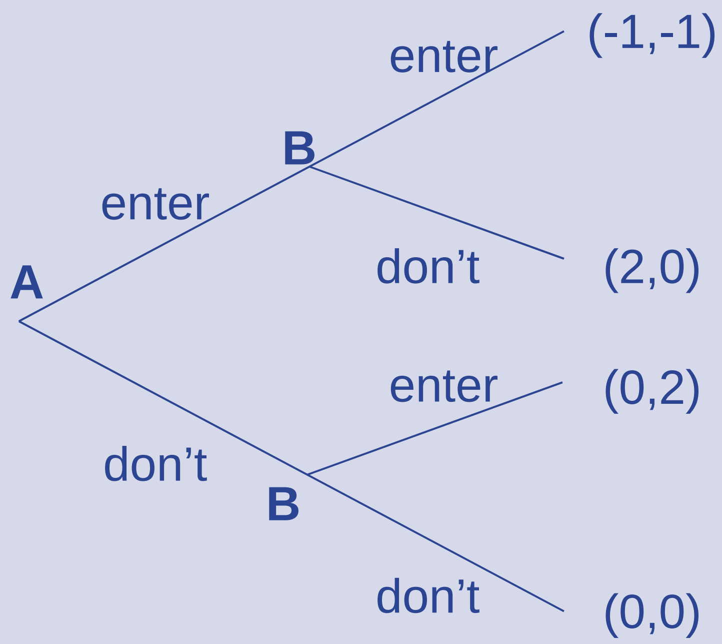
This game has two NE

For dynamic games, an extensive form representation is helpful.

Extensive Form

- Problem with the equilibrium (don’t, enter): It depends on the threat of B to enter if firm A has chosen to enter in the first stage.
- This threat is non-credible since firm B prefers to stay out given that firm A has entered.
- However, on the equilibrium path, i.e., when firm A stays out, firm B does not need to carry out its threat.
- Hence, (don’t, enter) is a Nash equilibrium.
- Can we refine the Nash equilibrium concept to rule out such implausible equilibria?

Extensive Form Representation

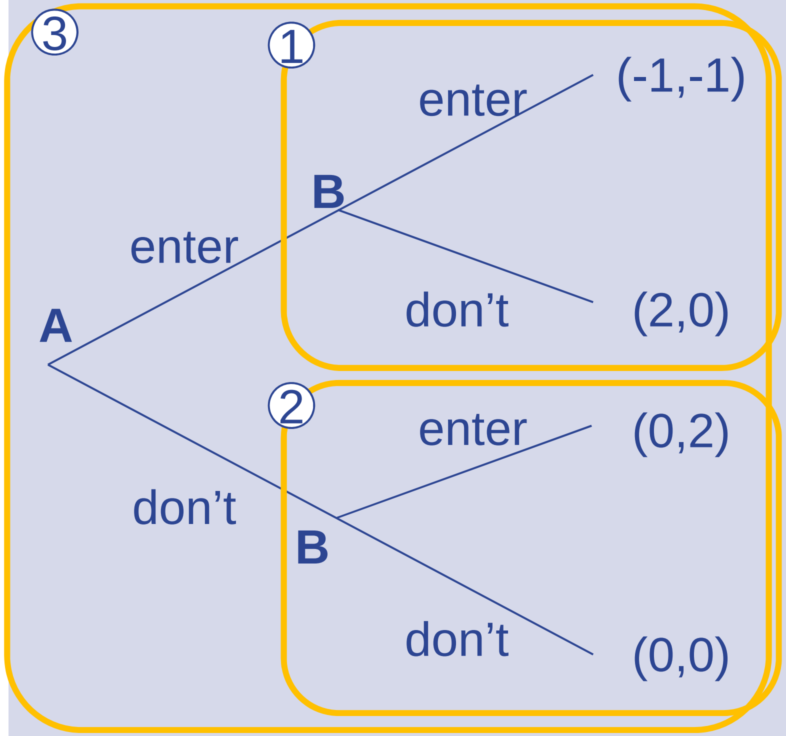


(don’t of firm A, enter of firm B) involves a non-credible threat

The way to solve sequential games is to identify subgame perfect equilibria.

Subgame Perfect Equilibrium

Extensive Form Representation



- Each decision node constitutes a "subgame".
- Our market entry game has 3 subgames.

Subgame Perfect Equilibrium

In a subgame perfect equilibrium (SPE), a strategy induces a NE in every subgame.

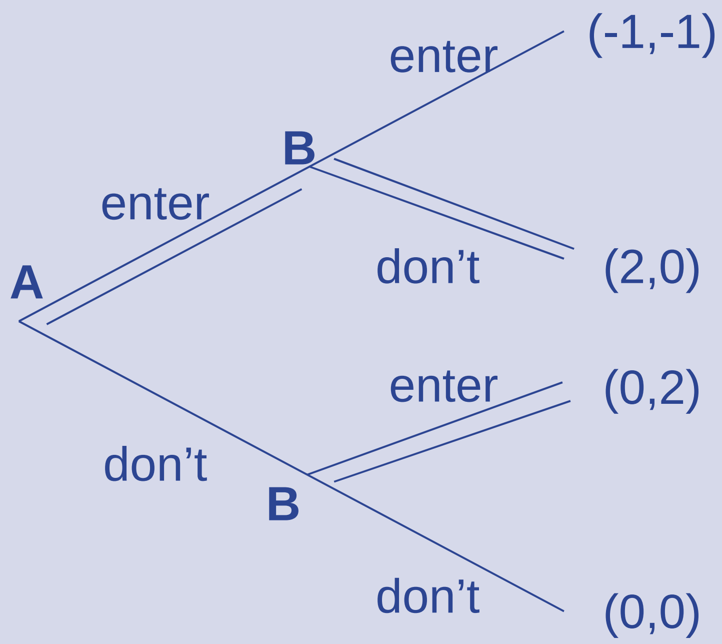
- Often we can find the SPE by applying "backward induction".
- Let us now apply backward induction.

Backward Induction Procedure

Extensive Form

- In subgame 1, (don’t) is the optimal choice of firm B.
- In subgame 2, (enter) is the optimal choice of firm B.
- Since firm A knows this, it knows that the resulting payoff from (enter) is 2 and the resulting payoff from (don’t) is 0.
- Therefore, in subgame 3, (enter) is the optimal choice of firm A.
- The Nash equilibrium (don’t of A, enter of B) is not subgame perfect because it relies on the non-credible threat that firm B enters in the second stage.

Extensive Form Representation



(enter of firm A, don’t enter of firm B) is the only subgame perfect eq.

Sequential game with continuous actions

Sequential Advertising Game

Backward Induction

- Consider the advertising game with continuous actions as above.
- Suppose now that firm 1 moves first and firm 2 follows after observing firm 1's advertising level.
- Profit of firm 2 is

$$\pi_2 = 1 + a_2 - \frac{1}{2}a_1 - a_2 a_1 - a_2^2 \rightarrow \max_{a_2}!$$

leading to the same reaction function as above:

$$a_2 = \frac{1 - a_1}{2}$$

- This is now a "true" reaction function since firm 2 observes a_1 and reacts to it.

Sequential game with continuous actions

Sequential Advertising Game

First Stage

- Firm 1 anticipates the optimal reaction of firm 2.

$$a_2 = \frac{1-a_1}{2}$$

- Firm 1 maximizes

$$\pi_1 = 1 + a_1 - \frac{1}{2}\left(\frac{1-a_1}{2}\right) - a_1\left(\frac{1-a_1}{2}\right) - a_1^2 \to \max_{a_1}!$$

- Solution: a_1*=3/4, a_2*=1/8
- Firm 1 produces more than in the simultaneous game whareas firm 2 produces less.

The value of commitment

Which firm benefits from sequential competition?

- Profit of firm 1 is 33/32; profit of firm 2 is 41/64.
- Therefore, firm 1 makes a higher profit compared to firm 2.
- Firm 1 also benefits relative to the static game, whereas firm 2 loses!
- Reason: Firm 1 can commit not to change its advertising level after stage 1 and can thereby influence firm 2's choice to its advantage.
- Interestingly, firm 1 does not play its best response (but firm 2 does)!

Limiting one self choices can be profitable in situations with strategic interaction

This has widespread applicability beyond economics,
e.g., negotiations or war situations.

Auctions are used in may economic contexts; we can analyze them as games.

Example: Sotheby’s

Situation

- You are an art trader. Your customer has sent you to an art auction at Sotheby’s to bid for a Van Gogh painting. Your customer has set a maximum price of X m. €.
- Of every Euro you save her, you can keep 50 €-Cent. If you don’t get the painting, you get nothing.
- You know that all other bidders are art traders with a similar arrangement with their customers.
- But you do not know the “stop out” price of the other traders.
- What is the optimal bidding strategy?

How do you bid optimally at ebay?

Optimal Bidding at

Bidding Rules

- The current bid amount shown on an auction listing is the *lowest amount* necessary for the high bidder to have outbid the next highest bidder's maximum by one bid increment, regardless of the high bidder's actual bid.
- When a new bidder bids on an auction, one of two things can happen:
- If the new bidder's maximum bid is *higher* than the old winning bidder's maximum bid, the previous winning bidder is outbid and the current bid amount is set by eBay to the previous winning bidder's maximum bid *plus* one bid increment.
- If the new bidder's maximum bid is *lower* than the currently winning bidder's maximum bid, the currently winning bidder remains the high bidder, and the current bid amount is set by eBay to the new bidder's maximum bid *plus* one bid increment.
- When the auction closes (i.e., when time runs out), the winning bidder wins the auction at whatever the current bid amount is.
- eBay offers an automatic system called *proxy bidding*, in which a bidder announces the maximum bid, and then eBay bids on behalf of the bidder and increases the bid incrementally until the maximum is reached.

Examples of the Bidding Rules

Optimal Bidding at ebay

Example

- Suppose that the bid increment is 0.50€.
- In addition, suppose the highest bid is 15.00€ and the second-highest bid is 10.00€.
- When a new bidder bids on an auction, the bidder sees the second-highest bid plus the increment, which is 10.50€.
- Now one of two things can happen:

  Since the bidder does not know the highest bid, she may bid 17.00€. The listing price shown at ebay then jumps up to 15.50€. The new bidder then knows that she outbid the previously highest bid.

  However, the new bidder may also bid 12.00€. The listing price shown at ebay then jumps to 12.50€. This shows the new bidder that she did not submit the highest bid.

This situation seems to be quite complex – but we can simplify it in form of a second price auction.

Second Price Auction

Game

- There are n-players, each player has an (independent private) valuation for a single, indivisible object: $x_1,\dots,x_n$
- All players simultaneously submit a bid $b_1,\dots,b_n$ to the auctioneer.
- The auctioneer will give the object to the one with the highest bid.
- The winning bidder has to pay the second highest bid, all other bidders get nothing and pay nothing.

- This game is strategically equivalent to our Sotheby's auction and ebay
- Why? When the auction starts, you would always only incrementally increase your bid, until you are the only bidder still active in the auction, i.e., have made a bid just a bit higher than the second highest bidder!

The second price, sealed bid auction has a unique equilibrium.

Solving the second price auction

Claim

- There is a unique equilibrium of the second price auction with independent private values. In the equilibrium, each bidder bids her true value: $b_i^* = x_i$

Proof

- $b_i < x_i$ **can never be optimal**, since increasing b_i to x_i increases expected profits: if $b_i > b^+$, nothing changes; if $x_i < b^+$, again nothing changes; **but** if $x_i > b^+ > b_i$ the bidder does not win but could make a strictly positive profit by bidding $b_i = x_i$.
- $b_i > x_i$ **can never be optimal**, since decreasing b_i to x_i increases expected profits: if b_i is not the highest bid, nothing changes, if b_i is the highest bid and the highest competing bid b^+ is below x_i, nothing changes, **but** if b^+ is between b_i and x_i, $b_i > b^+ > x_i$, the bidder avoids a loss of $b^+ - x_i$.
- Thus, increasing the bid above x_i, and decreasing it below x_i is not profitable.

Key Points

Summary

1. Translating strategic interaction into a game helps to analyze the underlying situation to gain predictions for outcomes and to ask the right questions.
2. In a Nash equilibrium, everybody behaves optimally given the choices of others.
3. A subgame perfect equilibrium selects plausible NE in dynamic games. Often, we can find them by using backwards induction.

Markets, Incentives and Ethical Management

3. Oligopoly

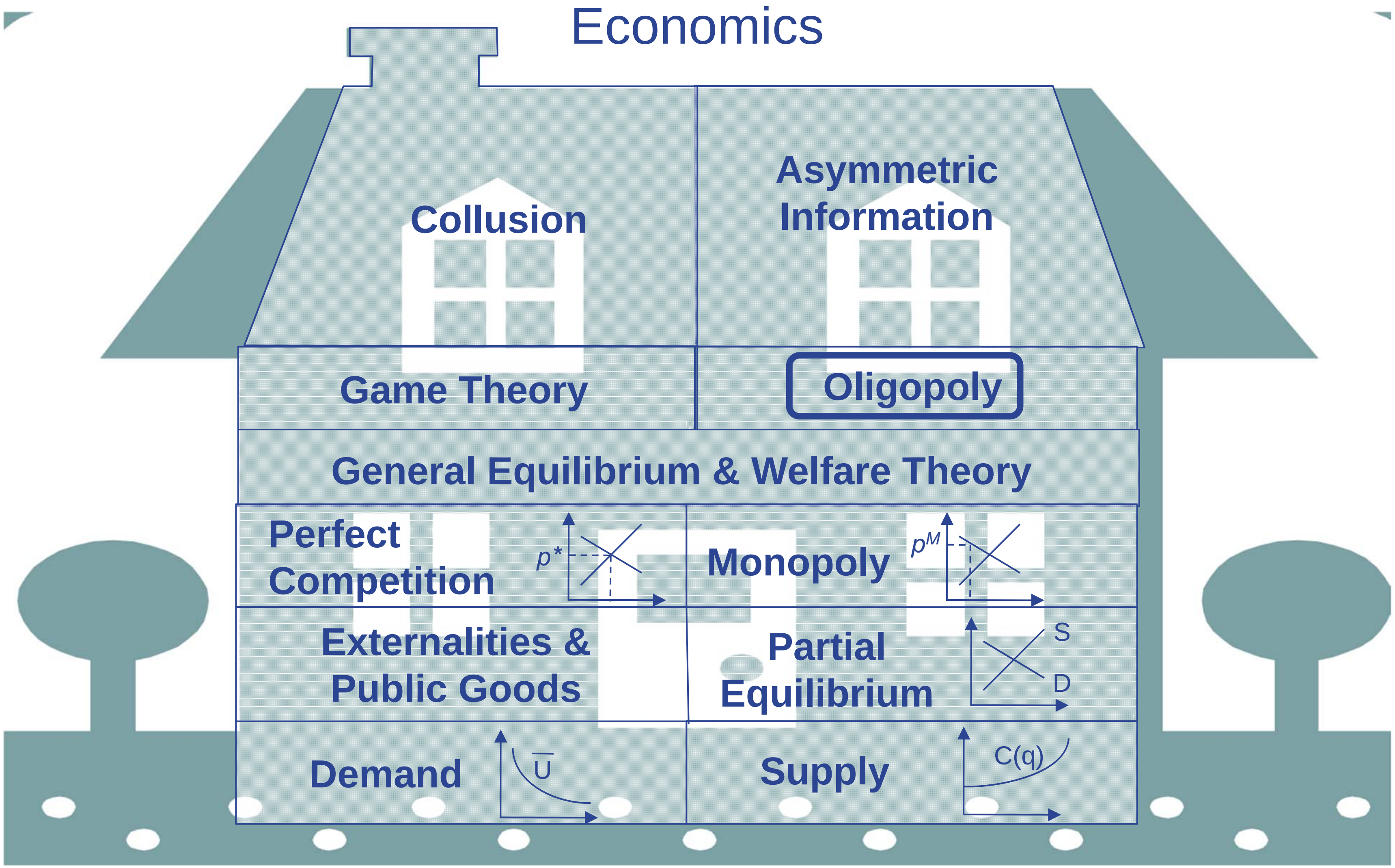
Prof. Dr. Markus Reisinger

Professor for Industrial Organization and Microeconomics

Frankfurt School of Finance & Management

m.reisinger@fs.de

Economics comes from Greek "oikos" which means "household".



Most firms interact strategically in oligopolistic markets.

Strategic Interaction

- Pure monopoly as well as perfect competition are rare.
- Most firms have some degree of market power but need to take rival’s behavior into account.
- Firms’ products can be homogeneous or differentiated

“Homogeneous Products”

- We focus on “homogeneous products” with a uniform price
- Examples:
  – Consumer products: Salt, flour, gasoline, loans
  – Business-to-business: oil, steel, grain

Aim of today's lecture

Aims

1. Modeling of strategic interaction
2. Understanding the force of pure price competition
3. Learning about the effect of the number of firms on the market outcome
4. Introducing an example of sequential competition

1	Market forms and strategic interaction
2	Bertrand competition
3	Cournot competition
4	Sequential quantity competition

Oligopoly is in-between the market forms we looked at so far.

Demand Side

- We always assumed that there are many buyers, so that they take the price as given.
- What does this mean? Recall the consumer problem: consumers maximize their utilities for given prices by adjusting the quantity they buy – they do not bargain over the price with the seller!
- In the following, we will maintain the assumption that consumers are price takers.

Supply Side

- In **competitive markets**, we assumed that there are many firms, so that each individual firm is too small to affect the price (price taking behavior / no market power assumption).
- In **monopoly**, we assumed that there is only a single firm, implying a lot of market power: the firm actually sets the price.
- Now we consider the intermediate case: “few” firms -> **Oligopoly**

Market forms can be characterized with respect to the number of players on each market side.

Market Forms		Sellers		
		One	Few	Many
Buyers	Many	**Monopoly**	**Oligopoly**	**Perfect Competition**
	Few			Oligopsony (e.g. grocery)
	One	Bilateral bargaining		Monopsony (e.g. labor market)

Conceptually, we can revert the roles of buyers and sellers to gain different forms of "buyer power" from none (perfect competition) to maximum buyer power (monopsony).

Monopoly and perfect competition are extreme market forms; most common is oligopoly.

Market Form	Monopoly	Oligopoly	Perfect Competition
Example	Patent protected drugs; (Windows OS in the 90s had over 90 % market share)	Most markets	New York Stock Exchange
Characteristics	Pharmaceutical firm can determine the market price for the drug; if the firm increases the price for the drug, it will receive a smaller demand	Firms can affect the market price but cannot determine the price alone since customers could go to competitors; reducing quantities might increase the market price nevertheless	Sellers do not set prices but make offers, i.e., quantities, conditional on the market price; reduction/expansion of offered quantities does usually not affect the price

In oligopoly, there is strategic interaction between firms.

Strategic Aspects of Oligopoly

- How will customers react to a strategic decision (e.g., pricing decision), given that they can switch between suppliers?
- How will competitors react to a strategic decision?

- The latter aspect is the core of strategic interaction:

  “Try to anticipate what competitors will do!”

Strategic interaction and the equilibrium concept apply to all forms of strategic choices.

Strategic Variables

- We call “**strategic choices**” also “**strategic variables**”
- Examples for strategic variables:
  - **Prices**: Each firm sets a price for its own product, taking into account that consumers may turn to competitors if prices are lower there
  - **Quantities**: Firms offer quantities and a market maker sets the price but firms know that they can affect the market price through their offered quantities (examples: small exchanges like the European Energy Exchange EEX with few suppliers, electricity market)
  - **Product differentiation**: Each firm offers the product in a different form or quality
    - “Horizontal product differentiation”: offering different styles of cars (BMW vs. Mercedes)
    - “Vertical product differentiation”: offering different quality levels (iPhone 11 Pro Max vs. iPhone 8)
  - **Location**: Firms offer the same product but at different locations (e.g., fuel stations at motorways or in cities)
  - **Information**: Firms might voluntarily reveal information about the quality of their product (e.g., ingredients, eco-labels etc.)

We will focus on specific forms of strategic interaction in the remainder of this course.

Strategic Interaction		Timing		
		Simultaneous Moves	Sequential Moves	Repeated Interaction
Strategic Variable	Price	**Bertrand**		**Models of collusion**
	Quantities	**Cournot**	**Stackelberg**	**Models of collusion**
	Product differen-tiation			

Focus of this chapter

Focus of next chapter

1	Market forms and strategic interaction
2	**Bertrand competition**
3	Cournot competition
4	Sequential quantity competition

Price competition with homogenous products

The Bertrand model

Assump-tions

- Consider a single market for a homogenous product.
- There are two firms, each produces a homogeneous product at constant marginal cost of c: $C_1 = c\,q_1$, $C_2 = c\,q_2$.
- Firms simultaneously set prices.
- Consumers will all buy from the firm with the lower price and firms face no capacity constraints (i.e., each firm is able to satisfy the total market demand $D(p)$ at any price); if both firms charge the same price, the demand is split evenly between the two firms.

What is the Nash equilibrium of this game?

Equilibrium prices when goods are homogeneous

The Bertrand model

Claim: The only equilibrium is one, in which firms charge $p_1 = p_2 = c$.

- Why is the claimed result an equilibrium?[1)]
  - If $p_1 = c$, choosing $p_2 = c$ is a best response, since choosing $p_2 > c$ would imply no sales for firm 2, i.e., also zero profits and no improvement versus $p_2 = c$.
  - Charging $p_2 < c$ would give firm 2 the whole demand, but firm 2 would be strictly worse off since it makes losses now.

- Why is there is no other equilibrium?
  - If $p_1 = p_2 = p > c$, both firms would make positive profits of

$$\frac{1}{2}D(p)(p-c)$$

  - Now consider a slight decrease of p_2 to $p_2 = p - \varepsilon$. Then, all consumers would go to firm 2, and firm 2 would make profits of

$$D(p-\varepsilon)(p-\varepsilon-c)$$

1) All arguments also apply when changing the roles of firm 1 and firm 2.

Equilibrium prices when goods are homogeneous

- A slight decrease of *p* by *ε* and monopolizing the market is always profitable if

$$\frac{1}{2}D(p)(p-c) < D(p-\varepsilon)(p-\varepsilon-c)$$

which can be rearranged to

$$\frac{D(p)}{2D(p-\varepsilon)} < \frac{p-\varepsilon-c}{p-c}$$

- One can see that for *ε->0,* the left hand side goes to ½, while the right hand side goes to 1 and it follows that for *ε* sufficiently small, undercutting the rival's price by *ε* is always profitable!

➢ More intuitively: instead of sharing the market with the other firm, each firm has always the incentive to slightly undercut *p* and gain the total market.

➢ By the same argument, $c < p_1 < p_2$ cannot be an equilibrium.

Bertrand paradox

The Bertrand model

Solution

- In the only equilibrium, prices are $p_1 = p_2 = c$.
- Thus, the same solution as under perfect competition arises (price = marginal cost).
- This is called "Bertrand Paradox", since it states the "already two firms are enough for achieving perfect competition", which is unrealistic.

Summary: If a market is characterized by the assumptions of the Bertrand model, we would expect quite fierce competition and low profit margins.

▶ Firms have strong incentives to differentiate their products to dampen competition.

Bertrand competition shows an extreme form of price competition.

Interpretation

- The Bertrand model is „good“ in so far as it makes realistic behavioral assumptions:
  - Firms set prices;
  - Consumers buy from the cheapest supplier.
- The Bertrand model is „bad“ as it yields a somewhat unrealistic result.

- However, it might be useful to look for a model with more realistic forecasts.
- What would such forecasts?
  - Oigopoly should leave profits to firms!
  - The number of firms should matter for the intensity of competition! (Recall that we considered this argument when introducing strategic interaction.)

1	Market forms and strategic interaction
2	Bertrand competition
3	Cournot competition
4	Sequential quantity competition

Cournot competition is competition in quantities.

Cournot Duopoly

Assump-tions

- Consider a single market for a homogenous product.
- There are two firms, each produces a homogeneous product at constant marginal cost of c: $C_1 = c\, q_1$, $C_2 = c\, q_2$, where $c < 1$.

- Demand function D describes the quantity that consumers buy as a function of the market price $D = 1 – p$.
- Supply Q is the aggregate quantity, that is, $Q=q_1 + q_2$.
- In equilibrium, D=Q, which implies Q=1-p

- Inverting the demand function gives the market price as a function of the quantity demanded $p = 1 – q_1 – q_2 \Rightarrow p = 1 – Q$.
- Firms choose simultaneously quantities q_1 and q_2, and then the market clearing price evolves.

Maximization of the profit function provides the "best response" function for each firm.

Solving the Cournot Duopoly

- Each firm maximizes its profit function (consider firm 1, firm 2 is similar):

$$\pi_1 = q_1(p-c) = q_1(1-q_1-q_2-c) \to \max_{q_1}! \qquad (1)$$

➢ Note that firms do not set prices here – their strategic variable is their quantity!

➢ In the end, since products are homogeneous, the price will always be the same for the two firms, depending on the sum of the quantities they offer.

- Taking the derivative of (1) wrt q_1, we derive the first order condition for (1):

$$(2) \qquad \frac{\partial \pi_1}{\partial q_1} = 1 - q_1 - q_2 - c - q_1 = 0$$

- Solving (2) for q_1 gives firm 1's reaction function which is the optimum choice of q_1 for any given level of q_2:

$$q_1 = \frac{1-q_2-c}{2}$$

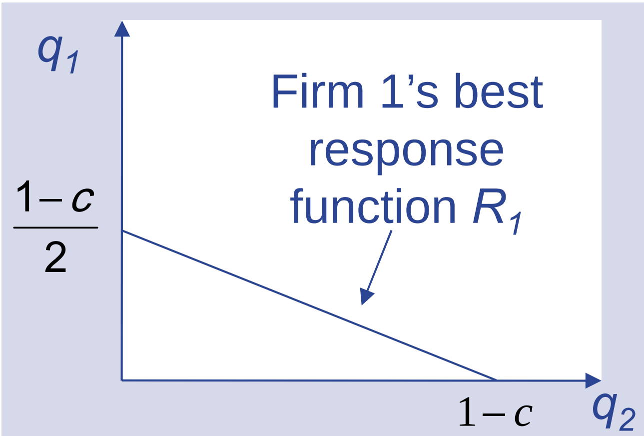
Maximization of the profit function provides the "best response" function for each firm.

Solving the Cournot Duopoly (cont.)

- The best response function of firm 1 (R_1) assigns a quantity q_1 to every given quantity of firm 2 (q_2) that maximizes firm 1's profit

$$R_1 \equiv q_1 = \frac{1 - q_2 - c}{2}$$

- Best-response function is falling; quantities are *strategic substitutes*



The Cournot equilibrium is given by the intersection of the best response functions.

Solving the Cournot Duopoly (cont.)

Due to the symmetry of the two firms, firm 2's reaction function is similar:

(3) $$q_2 = \frac{1 - q_1 - c}{2}$$

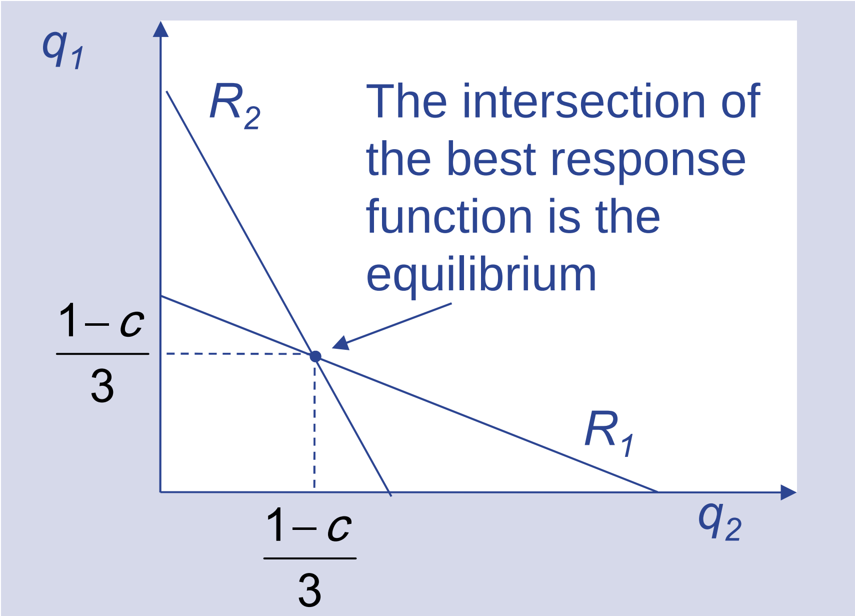
We can now solve equations (2) and (3) for q_1 and q_2.

This yields:

(4) $$q_1{}^* = q_2{}^* = \frac{1-c}{3} \Rightarrow$$

(5) $$p^* = \frac{1+2c}{3} > c \Rightarrow$$

(6) $$\pi_1{}^* = \pi_2{}^* = \frac{(1-c)^2}{9} > 0.$$



Additional insights can be gained from a symmetric *n*-firm Cournot oligopoly.

n-firm Cournot Oligopoly

Assump-tions

- Consider a market for a homogenous product.
- There are n, $n > 1$, firms, each firm produces the homogeneous product at constant marginal cost of c: $C_i = c\, q_i$, $i = 1,2,\ldots n$, where $c < 1$.
- The inverse market demand function is

$$p = 1 - \sum_{j=1}^{n} q_j$$

- Firms simultaneously choose quantities q_i , $i=1,2,\ldots,n,$ and then the market clearing price evolves.

Equilibrium of the symmetric *n*-firm Cournot oligopoly.

n-firm Cournot Oligopoly

Solution

- Each firm maximizes its profits wrt. the quantity:

$$\pi_i = q_i(p-c) = q_i\left(1-\sum_{j=1}^{n} q_j - c\right)$$

- Taking the derivative wrt. q_i gives the first-order condition:

$$\text{(7)} \qquad \frac{\partial \pi_i}{\partial q_i} = \left(1-\sum_{j=1}^{n} q_j - c\right) - q_i = 0$$

- Since all firms are symmetric we know that they all will choose the same q_i . Hence: $\sum_{j=1}^{n} q_j = nq_i$

➢ Inserting this into the last expression (7) gives:

$$(1-nq_i-c)-q_i = 0$$

Equilibrium of the symmetric *n*-firm Cournot oligopoly.

- Solving for the eq. variables yields:

$$q_i^* = \frac{1-c}{n+1}$$

$$p^* = 1 - nq^* = 1 - n\frac{1-c}{n+1} = \frac{n+1-n+nc}{n+1} = \frac{1+nc}{n+1} > c$$

$$\pi_i^* = q_i^*(p_i^* - c)^* = \left(\frac{1-c}{n+1}\right)\left(\frac{1+nc}{n+1} - c\right) = \left(\frac{1-c}{n+1}\right)\left(\frac{1+nc-nc-c}{n+1}\right) = \frac{(1-c)^2}{(n+1)^2}$$

- It is easily seen that profits decrease with the number of firms.
- Realistic outcome: an increasing number of firms exacerbates competition!

In Cournot competition, increasing the number of firms increases the level of competition.

Interpretation of the *n*-firm Cournot Oligopoly

- In Cournot competition, firms have market power, i.e., their strategic interaction leads to price above marginal cost.
- If we increase the number of firms, the “mark-up” over the marginal cost decreases and profits of each firm go down.
- In the limit, where n goes to infinity, we approach the competitive outcome (*price = marginal cost*, zero profits for the firms).
- This is intuitive: when the number of competitors increases, the level of competition should (gradually) increase and the ability to sustain high profit margins should (gradually) decrease.

The Cournot model yields a realistic theoretical forecast.

Interpretation

- The Cournot model is „good“ as it yields intuitive and realistic forecasts:
  - In oligopoly, firms (always) have market power to raise the market price above the competitive level.
  - However, the ability is weakened if one increases the number of firms.
- The Cournot model is analytically very tractable – thus it has become a workhorse model in oligopoly theory.
- An important drawback of the Cournot model is that it rests on unrealistic behavioral assumptions:
  - Firms do (usually) not just choose quantities, they rather set prices.
  - There is (usually) no “auctioneer” determining prices.

The Stackelberg model analyzes quantity competition with sequential moves.

Stackelberg Model

1. There are two symmetric firms.
2. Firms compete in quantities.
3. Firm 1 determines its quantity first.
4. Firm 2 can observe the quantity choice of firm 1 before making its own output decision. Firm 1 is committed to its quantity.

In the Stackelberg model, firm 1 can anticipate firm 2's reaction.

Stackelberg: Analytical Solution

- Consider the linear demand example of the Cournot duopoly, but assume firm 1 moves first.
- For any given choice q_1, firm 2's reaction is: $q_2 = \frac{1 - q_1 - c}{2}$
- Anticipating this, firm 1 maximizes:

$$\pi_1(q_1) = (p-c)q_1 = \left(1 - q_1 - \frac{1-q_1-c}{2} - c\right)q_1 \to \max_{q_1}$$

$$FOC: \frac{\partial \pi_1}{\partial q_1} = \frac{1-q_1-c}{2} - \frac{q_1}{2} = 0$$

$$q_1 = \frac{1-c}{2} \to q_2 = \frac{1-c}{4} \to p = \frac{1+3c}{4}.$$

- **Firm 1 produces a larger quantity than firm 2.**
- **Since the equilibrium price is the same for both firms, the profit of firm 1 is larger than the one of firm 2.**

Commitment benefits the first mover

Comparison with Cournot game

- Firm 1 produces a larger quantity than in the simultaneous game, i.e., $(1-c)/2 > (1-c)/3$.
- It also obtains a higher profit, i.e., $(1-c)^2/8 > (1-c)^2/9$.
- The opposite is true for firm 2.
- The intuition is the commitment effect, explained in the last chapter.

Firm 1 has a first-mover advantage.

We need models to understand strategic interaction on firm level and for the economic analysis.

Summary

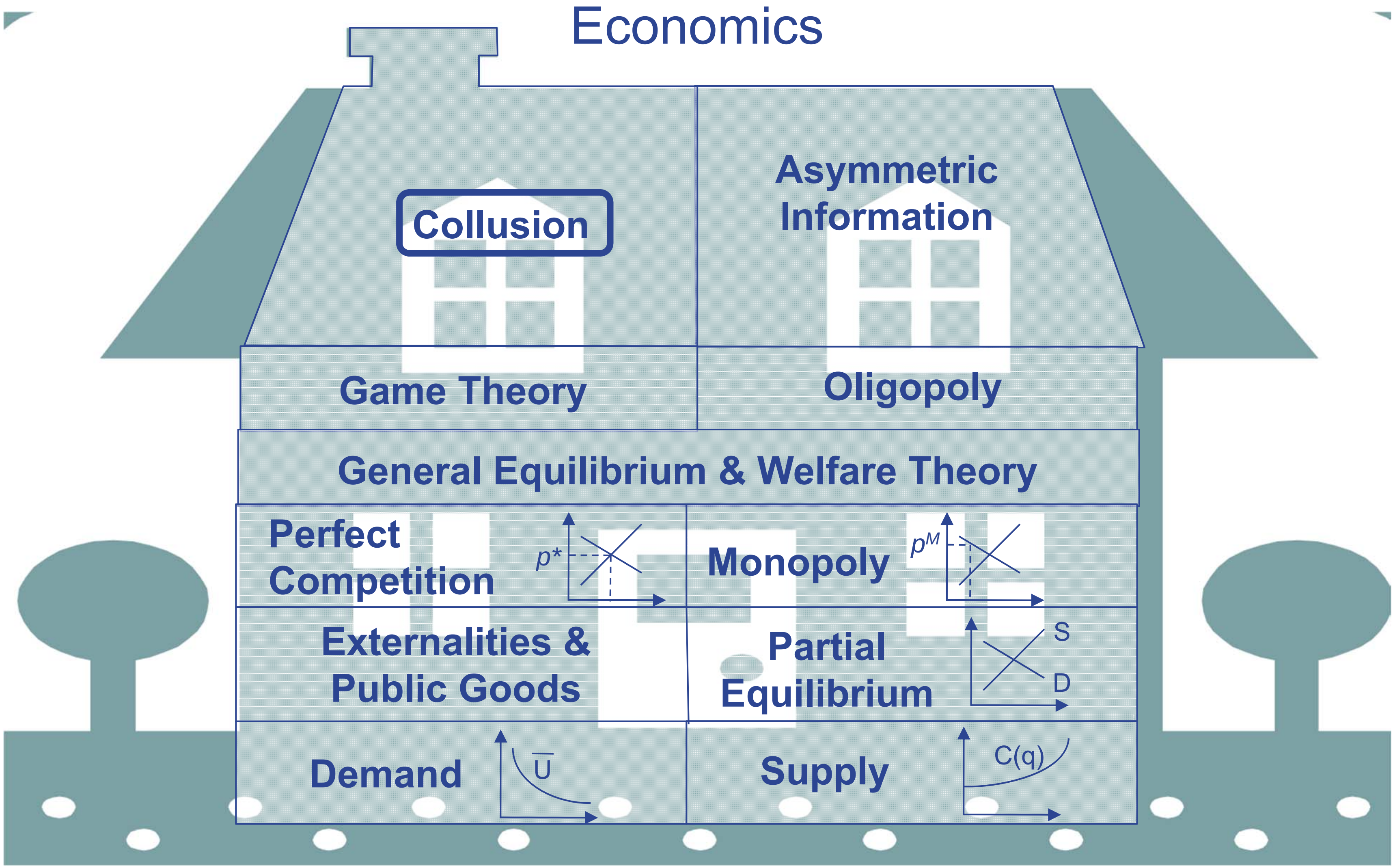
- Strategic interaction is important firms' decisions ("How will competitors react?"). If firms have market power, the outcome may differ from the perfectly competitive one.
- Important "workhorse" models for oligopolistic competition with homogenous goods are Bertrand competition and Cournot competition.
- Bertrand competition is competition in prices and implies an extremely high intensity of competition ("Bertrand Paradox").
- Cournot competition is competition in quantities and yields more realistic outcomes.
- With quantity competition, moving first provides a first-mover advantage (Stackelberg equilibrium).

Markets, Incentives and Ethical Management

4. Cartels and Tacit Collusion

Prof. Dr. Markus Reisinger
Professor for Industrial Organization and Microeconomics
Frankfurt School of Finance & Management
m.reisinger@fs.de

Economics comes from Greek “oikos” which means “household”.



1	**Explicit versus Implicit Cartels**
2	**Tacit Collusion**

An explicit cartel

The vitamin cartel

- As seen in previous lectures, price competition leads to low profits.
- Price-fixing agreements are desirable for firms but are forbidden by competition law.
- Cartels nevertheless form and last.

Example: The vitamin cartel

- Worldwide market for bulk vitamins is very valuable (EUR 3.25 billion in 1999).
- Highly concentrated market (Hoffmann-La Roche, BASF, and Aventis control between 65% and 90% of the market for different vitamins).
- Price fixing by these firms was proven by the European Commission.

Fines in the vitamin cartel

The Vitamin Cartel (cont.)

- The then-record fine of EUR 855.22 million on eight companies for price-fixing and market-sharing arrangements.
- Cartels for some vitamins endured from 1989 to 1999.
- In the U.S., investigation by the Department of Justice:

Hoffmann La-Roche pleaded guilty for worldwide conspiracy to raise and fix prices and allocate market shares.

Criminal fine: US dollar 500 million (BASF: 225 million).

“Most damaging series of cartels the Commission has ever investigated“
(Mario Monti, then European Competition Commissioner)

Other Cartels

Lysine Price-Fixing Cartel

• Cartel on animal feed additive lysine in the 1990s, involving American and Japanese firms.

• Cartels raised prices by more than 70%.

• First successful prosecution of a cartel by the Department of Justice in 40 years.

• Started by whistleblower Mark Whitacre (adapted into a movie with Matt Damon in 2009); basically the beginning of a radically new thinking of governments and antitrust authorities on cartels.

Fines: 3 top executives of Archer Daniels Midland (ADM) were sent to jail for three years (sentence in total 99 months).

Monatery fines: $100 Million in the US, $50 Million elsewhere

More than $500 Million in class action cases to plaintiffs and consumers.

Other Cartels

Euro Interest Rate Derivatives Cartel

- December 2016: European Commission fined Crédit Agricole, HSBC, and JPMorgan Chase a total fine of € 485 Million.
- Banks fixed Euro Interbank Offered Rate (EURIBOR) and/or the Euro Over-Night Index Average (EONIA) between 2005 and 2008.
- Statement of the Commission:

“The participating traders of the banks were in regular contact through corporate chat-rooms or instant messaging services. The traders' aim was to distort the normal course of pricing components for euro interest rate derivatives. They did this by telling each other their desired or intended EURIBOR submissions and by exchanging sensitive information on their trading positions or on their trading or pricing strategies.”

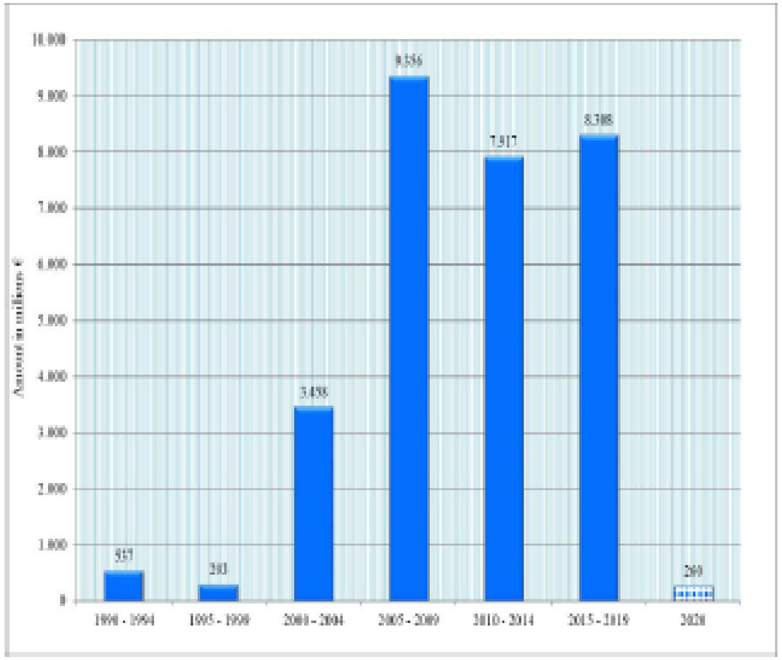
Competition policy

- Collusion and price-fixing is one of the most important issues in competition policy
- Mario Monti, former European Competition Commissioner (2000):
  “Cartels are cancers on the open market economy.”
- Nelly Kroes, former European Competition Commissioner (2009):
  “I don’t want to merely destabilize cartels. I want to tear the ground from under them.“

- Often, leniency programs are introduced offering reduced penalties to cartel members in exchange for cooperating with enforcement authorities.

Fines in Europe

Year	Amount in €*)
1990 - 1994	537 491 550
1995 - 1999	292 838 000
2000 – 2004	3 458 421 100
2005 – 2009	9 355 867 500
2010 – 2014	7 917 218 674
2015 – 2019	8 307 828 000
2020	260 443 000
total	**30 130 107 824**



Source: European Commission (July 14, 2020)

Size of Cartel Fines

Ten highest cartel fines per undertaking since 1969

Year	Undertaking**	Case	Amount in €*
2016	Daimler	Trucks	1 008 766 000
2016	DAF	Trucks	752 679 000
2008	Saint Gobain	Carglass	715 000 000
2012	Philips	TV and computer monitor tubes	705 296 000 of which 391 940 000 jointly and severally with LG Electronics
2012	LG Electronics	TV and computer monitor tubes	687 537 000 of which 391 940 000 jointly and severally with Philips
2016	Volvo/Renault Trucks	Trucks	670 448 000
2016	Iveco	Trucks	494 606 000
2013	Deutsche Bank	Euro interest rate derivatives (EIRD)	465 861 000
2001	F. Hoffmann-La Roche	Vitamins	462 000 000
2007	Siemens	Gas insulated switchgear	396 562 500

Source: European Commission

Cartel bloopers (I)

Feb 1982: Phone call between Robert Crandall (CEO, American Airlines) and Howard Putnam (CEO, Braniff Airlines)

Cartel bloopers (I)

- **Crandall**: I think it’s dumb as hell for Christ’s sake, all right, to sit here and pound the **** out of each other and neither one of us making a ****ing dime.

- **Putnam**: Do you have a suggestion for me?

- **Crandall**: Yes. I have a suggestion for you. Raise your goddamn fares twenty percent. I’ll raise mine the next morning. You’ll make more money and I will too.

- **Putnam**: We can’t talk about pricing.

- **Crandall**: Oh bull ****, Howard. We can talk about any goddamn thing we want to talk about.

Cartel bloopers (II)

- Hasbro (Office of Fair Trading UK, 2003)
- Toy manufacturer Hasbro organized a price fixing agreement between retailers Argos and Littlewoods with respect to Hasbro's products:
- Email from Hasbro sales director Mike Brighty to Neil Wilson and Ian Thomson (19 May 2000):

'Ian ... This is a great initiative that you and Neil have instigated!!!!!!!!! However, a word to the wise, never ever put anything in writing, its highly illegal and it could bite you right in the arse!!!! suggest you phone Lesley and tell her to trash? Talk to Dave. Mike'

Cartel bloopers (III)

Algorithmic Pricing

- In April 2011, the academic book "The Making of a Fly" by Peter A. Lawrence was sold online by two booksellers, Profnath and Bordeebook.
- The two algorithms were programmed to do the following:

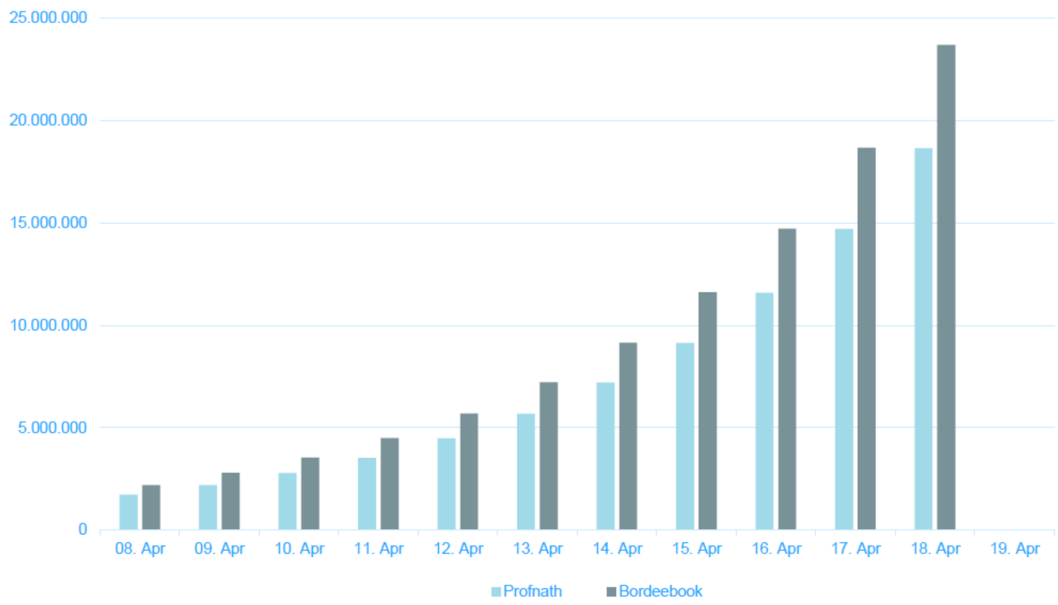
  Profnath: sell the book at 99.8% of Bordeebook's price

  Bordeebook: sell the book at 27% above Profnath's price

What do you think happened?

Cartel bloopers (III)

- Result: book was priced at more than $23m.
- Human intervention was required to bring the prices back down (to $106 and $135)



In general, pricing through algorithms can be problematic to prove collusion.

Implicit cartels

Implicit collusion

- Cartels can also be reached in a ‘tacit’ way.
- Implicit understanding that deviation from collusive tacit agreement will be met by some form of punishment.

$\longrightarrow$ “Meeting of the minds“

Edward Chamberlin (1933):

“If each (entrepreneur) seeks his maximum profit rationally and intelligently, he will realize that when there are only two or few sellers, his own move has a considerable effect upon his competitors, and that this makes it idle to suppose that they will accept without retaliation the losses he forces upon them. Since the result of a (price) cut by any one is inevitably to decrease his own profit, no one will cut and although the sellers are entirely independent, the equilibrium result is the same as though there were a monopolistic agreement between them.“

Implicit versus explicit cartels

Boundaries between lawful and unlawful collusion are blurry.

- Spectrum Auctions (Germany, 1999)
- German government auctioned off ten blocks of spectrum.
- Bidding rule: any bid must be at least 10% higher than the current highest bid

Mannesmann and T-Mobile were among the biggest bidders

- Mannesmann‘s initial bids:
  Blocks 1-5: 20 million DM/megahertz
  Blocks 6-10: 18.18 million DM/megahertz
- Why 18.18?

Implicit versus explicit cartels

Example continued:

- Adding 10% to 18.18 is 20.
- Was Mannesmann signaling to T-Mobile that each should win 5 blocks at 20 million?
- In the next round, T-Mobile bid 20 million on blocks 6-10. There were no subsequent bids.
- This is tacit collusion and it is lawful.
- If Mannesmann and T-Mobil had spoken to each other and exchanged assurances that each would buy 5 blocks for 20 million then that is explicit collusion and is unlawful.
- Effects are the same.

A (perhaps) implicit cartel

The NASDAQ

- Second largest stock market in the U.S. (after the NYSE).
- Online trade.
- Multiple market makers for each stock (on average between 10 and 20).
- Market maker posts two prices:

  “Ask”-price: Price at which he is willing to sell shares

  “Bid”-price: Price at which he is willing to buy shares
- These prices can only be in increments of an eighth of a dollar.
- Lowest ask-price and highest bid-price constitute the market prices.
- Market makers compete against each other.
- Higher bid-ask spreads imply higher profit for market makers.

Collusion between market makers?

The NASDAQ (cont.)

Towards the end of the Eighties and in the beginning of the Nineties, an unusually high percentage of all bid and ask prices where quoted in “even” heights of dollar.

Bid-ask spread was often at least 25 cents and sometimes 50 cent.
(In 1991: bid-ask spread was 1/8 in 10% of all cases, 2/8 in 39% of all cases, 3/8 in 5% of all cases, and 4/8 in 33% of all cases).

Daily trading volume: 650 million shares at NASDAQ

→ Spread increase by 1/8 increases the profit of market makers by U.S. Dollar 81.25 million.

Can this behavior of market makers be explained by implicit collusion?

1	Explicit versus Implicit Cartels
2	**Tacit Collusion**

Repeated Bertrand competition

Modeling framework

Firms

Two symmetric firms produce a homogeneous good with constant marginal costs equal to c. No fixed costs are considered.

Demand

Total demand is: $D(\min(p_1,p_2))$
Demand of firm 1, given p_1 and p_2 is:

$$D_1(p_1, p_2) = \begin{cases} D(p_1) & \text{if} \quad p_1 < p_2 \\ 1/2D(p_1) & \text{if} \quad p_1 = p_2 \\ 0 & \text{if} \quad p_1 > p_2 \end{cases}$$

Timing

Firms compete for T periods. No physical link between periods.
Firms choose prices simultaneously in each period.
Discount factor equals δ. δ close to 1 represents a high value for future profits.

Betrand competition with finite horizon

Modeling framework (cont.)

Profit function

Profit function of firm i in period t:

$$\pi_i^t = (p_i^t - c) D_i^t (p_1^t, p_2^t)$$

Solution concept

Subgame perfect Nash equilibrium:
Game is solved by backward induction.

Static game

Nash equilibrium of the static game:
Both firms set $p_1 = p_2 = c$.

Question

Can we find an equilibrium in which firms tacitly collude, that is, set the monopoly price in each period and obtain the monopoly profit?

Equilibrium with finite horizon

Backward induction

Period T:

- Each firm seeks to maximize its per-period profit as nothing ensues.
- The unique Nash equilibrium in period T is then: $p_1^T = p_2^T = c$

Period T-1:

- Price choices in period T do not depend on what happened in period T-1.
- Hence, firms compete in T-1 as if it were the last period.
- Again, the unique Nash equilibrium in period T-1 is: $p_1^{T-1} = p_2^{T-1} = c$

Periods before T-1:

- We can continue to solve the game like this in each period.
- Same argument carries on up to the first period.

Result with finite horizon is not convincing

Discussion

Summary:

With finite horizon tacit collusion cannot emerge!

Result is not convincing:

- If T is large, we would probably expect that firms behave cooperatively for some time and non-cooperatively towards the end. (This is observed in laboratory experiments.)
- Problem: last period effect.

Question:

Can the last period effect be avoided when allowing for infinitely many periods?

Bertrand competition with infinite horizon

Infinitely many periods

Interpretation: There is no known end date to the game. In each period, there is a strictly positive probability that firms also compete in the next period. (This probability can e.g., be incorporated in δ.)

Note: Competitive behavior like in the finite Bertrand game

$$p_1^t = p_2^t = c \quad \forall \quad t = 1,\dots,T,$$

is still a subgame perfect equilibrium in the infinitely repeated game.

If the opponent chooses p=c in every period, it does not pay off to raise the price above c.

BUT: This is no longer the unique subgame perfect equilibrium.

Grim trigger strategies

Collusive strategy for firm i, i=1,2.

In period 1:

Start by choosing the monopoly price, i.e., $p_i^1 = p^M$

In periods t=2,3,…:

Keep choosing the monopoly price as long as both firms have done so in all previous periods.

Cooperation phase

If one firm has deviated from the monopoly price, choose price equal to marginal costs (static Nash equilibrium).

Punishment phase

Strategy is called grim trigger strategy because deviation from collusive behavior triggers punishment and firms stick to non-cooperative behavior forever. (Note that the punishment constitutes a Nash equilibrium.)

Subgame-perfect equilibrium

Profit with collusion

We will now determine under which conditions, the grim trigger strategy of the previous slide constitutes a subgame perfect Nash equilbrium.

Since there is no terminal period, we cannot apply backward induction.

Way to proceed: Suppose firm 2 follows the grim trigger strategy. What is the best response for firm 1?

Payoff of firm 1 when sticking to the grim trigger strategy in period 1 (i.e., the profit of firm 1 in collusion):

$$\pi_1 = \frac{\pi^M}{2}\left(1+\delta+\delta^2+\delta^3+\ldots\right) = \frac{\pi^M}{2}\frac{1}{1-\delta}.$$

where $\pi^M = (p^M - c)D(p^M).$

Determining the critical discount factor

Deviation strategy

If firm 1 deviates in period 1 and slightly undercuts the monopoly price of firm 2, its payoff is:

$$\pi_1 = \pi^M + \delta \cdot 0 + \delta^2 \cdot 0 + \ldots = \pi^M .$$

Thus, a deviation from cooperative behavior in the first period does not pay off if and only if:

$$\frac{\pi^M}{2} \frac{1}{1-\delta} \geq \pi^M \qquad \Leftrightarrow \qquad 1 \geq 2 - 2\delta$$

or

$$\delta \geq \frac{1}{2}.$$

Critical discount factor

The trigger strategies constitute a subgame perfect equilibrium of the infinitely repeated Bertrand game if and only if the discount factor δ is larger than a threshold value δ_{TS}=1/2, that is,

$$\frac{1}{2} \leq \delta < 1.$$

Intuition: Firms must put sufficient weight on future losses to offset the temptation to secure immediate gains by deviating.

Tacit Collusion can be sustained by grim trigger

Deviation strategy (cont.)

Consider next firm 1's decision whether or not to deviate in period t, t=2,3,..:

• If no firm has deviated from its strategy so far, the same reasoning as above applies, i.e., firm 1 does not gain from deviating in this period either.
• If a firm deviated in period t-1, it is optimal for firm 1 to choose p=c, since the rival does the same.

The same arguments apply for firm 2.

When competition is repeated over an infinite horizon, tacit collusion can be sustained by grim trigger strategies as long as firms have a large enough discount factor.

Competition between N firms

Critical discount factor with N firms

Suppose now that N firms compete instead of only 2:

Deviation profit is unchanged because the undercutting firm obtains the monopoly profit.

Punishment profit is also the same due to price equals marginal costs.

Collusion profit now needs to be shared between N firms.

Hence, the formula becomes:

$$\frac{\pi^M}{N}\frac{1}{1-\delta} \geq \pi^M \quad \Leftrightarrow \quad \delta \geq 1 - \frac{1}{N}$$

Critical discount factor is increasing in N. Collusion is more difficult to sustain, the larger is the number of firms.

Collusion is more likely in markets, in which a small number of firms is active.

Analysis for general competition models

General analysis

Analysis so far considered Bertrand competiton. However, a similar analysis can be applied for many modes of competition.

- If a firm cooperates, it obtains (as long as all other firms cooperate as well):

$$\frac{\pi^M}{N} + \delta \cdot \frac{\pi^M}{N} + \delta^2 \cdot \frac{\pi^M}{N} + \ldots = \frac{\pi^M}{N} \frac{1}{1-\delta}$$

- Denote the deviation profit by π^D and the punishment profit by π^P. Then, by deviating the firm obtains:

$$\pi^D + \delta \cdot \pi^P + \delta^2 \cdot \pi^P + \ldots = \pi^D + \delta \pi^P \frac{1}{1-\delta}$$

A general formula

General critical discount factor

- A firm prefers to cooperate if and only if

$$\frac{\pi^M}{N}\frac{1}{1-\delta} \geq \pi^D + \delta\pi^P \frac{1}{1-\delta}$$

- Solving this for δ yields:

$$\delta \geq \frac{\pi^D - \pi^M / N}{\pi^D - \pi^P}.$$

- In the Bertrand example $\pi^P = 0, \pi^D = \pi^M, N = 2$ implying that the critical discount factor equals ½.

(Note that it is also possible to sustain cooperative profit levels between the monopoly profit and the punishment profit as a subgame perfect Nash equilibrium.)

Can tacit collusion explain the NASDAQ example?

NASDAQ

- Market makers interact frequently and over long periods with the same market makers.

  Market makers essentially play an infinitely repeated game

- Market makers trade every day.

  Discount factor is very high

- NASDAQ allows brokers to direct an order to a particular market maker, even if this dealer is not quoting the best price, as long as the dealer has agreed to match the best price available.

  Market maker gets not necessarily more consumers by undercutting. This makes it less attractive to deviate from collusive prices.

Evidence for Collusion at NASDAQ

Explicit Collusion

These factors all point to the case that there was indeed collusion going on between market makers.

Department of Justice indeed found evidence for collusion. However, argument relied on explicit collusion. They had found tapes of phone conversations between dealers in which they arranged collusive pricing.

- NASDAQ made a number of changes:

  Investors can now compete directly with market makers.

  Limit order display rule: Market makers are forced to display all orders which quote better prices than those offered by market makers.

Summary

Summary

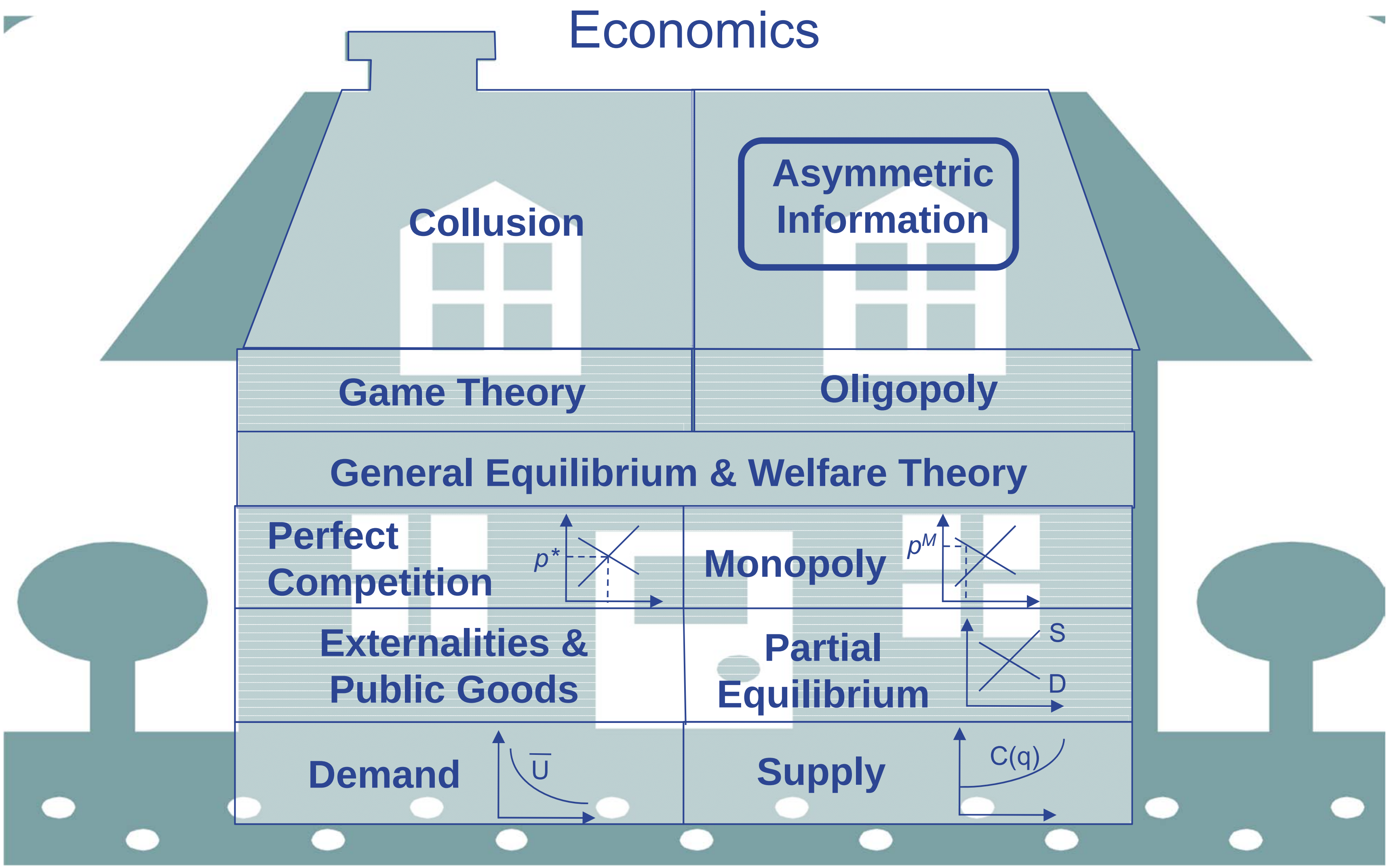
- Cartels are profitable for firms but they can lead to large damages for society.
- In a finitely repeated game, collusion cannot be sustained by firms due to the last period effect.
- In an infinitely repeated game, collusion can be sustained by simple grim trigger strategies if the discount factor is large enough.
- Collusion is harder to sustain the more firms are involved in the implicit cartel.

Markets, Incentives and Ethical Management

5. Asymmetric Information

Prof. Dr. Markus Reisinger
Professor for Industrial Organization and Microeconomics
Frankfurt School of Finance & Management
m.reisinger@fs.de

Economics comes from Greek “oikos” which means “household”.



Asymmetric information is a core topic in economics and in business reality.

Asymmetric Information

- One party knows something, the other party does not know.
- One party does something which cannot be observed by the other party.
- (Both parties have the same information but this information cannot be verified by courts).

Example

- Job interview: the employer asks: “Are you willing to work hard and carry on learning when employed?” You answer “Yes!” – but can the employer trust you?
- CEO compensation: How can the shareholders ensure that the CEO acts in the interest of the shareholders?

Aim of today’s lecture

Aims

1. What is “asymmetric information”?
2. Learn about “hidden information / adverse selection” problems.
3. Learn about “signaling”, which is a potential solution to adverse selection.
4. Learn about “hidden action / moral hazard” problems.

1	What is asymmetric information?
2	Hidden Information: Adverse Selection
3	Hidden Information: Signaling
4	Hidden Action: Moral Hazard

The “market for lemons” is the classic example for asymmetric information.

Example 1: Market for Used Cars

Situation

- You want to buy a used car.
- You are not a car expert, i.e., you cannot very well distinguish whether the used car offered by the car dealer is in good condition or not (is a “lemon”).
- You know that the car dealer knows more about the quality of the car and that the car dealer knows that you are not a car expert.
- The quality of the car cannot (ex post, or at a reasonable price ex ante) be verified.
- How does this influence the market outcome?

- A related question is whether the market outcome can be improved in case the quality can be verified ex post but only at high costs.

A second example are cases with asymmetric information on the “effort level” of one party.

Example 2: Labor Contracts

Situation

- You work as a consultant on a case with a client.
- The partner you are working for shows up only at the final presentation, she does not / cannot control how hard and good you actually worked on the case.
- The partner will see whether the client is satisfied with your work or not, but the level of satisfaction is influenced by many factors you cannot influence and which cannot be observed by your partner.
- Your partner wants to induce you to work hard – but how can she do so if she cannot directly control your effort level and cannot directly infer it from the outcome of the project (level of satisfaction of the client)?

It is important when the asymmetric information arises to conceptually understand the problem.

When does asymmetric information arise?

time

Before

The car is high- or low-quality.

Interaction/Contract

- Trading a car
- Employment contract

After

Consultant works hard or shirks.

- We think of different “types” of sellers: high types and low types
- The uninformed party knows the distribution of types but not the true type of the actual counterpart.
- What is the optimal strategy of buyers?

- We look for the optimal contract, which the uninformed party (i.e., the partner) can offer.
- What is the optimal contract that the partner (principal) proposes to the consultant (agent)?

We specify our used car market example.

Example: Used Car Market

Situation

- There are 3 possible types of car dealers:
  - high types (H): sell good cars which are worth 12 to a buyer and incur costs of 9.
  - medium types (M): sell medium cars worth 6 to a buyer and incur costs of 4.
  - low types (L): sell low value cars worth 1 to a buyer and incur costs of 0.
  - All types are equally likely.
  - A risk neutral buyer wants to buy a car. The buyer cannot infer the type of the seller and cannot make any ex post claims if the value of the car is lower than expected or promised by the seller.
  - The seller has all bargaining power, i.e., makes a “take-it-or-leave-it” offer to the buyer.
  - What is the market outcome?

As always, we first need to pin down the players objective functions.

Objective Functions

- **Buyer** wants to maximize expected gains from trade: $U^B = E[v_i] - p$, where $i = \{H,M,L\}$, and p is the price demanded by the seller.

Type	Probability	Value v	Cost c
H	1/3	12	9
M	1/3	6	4
L	1/3	1	0

- **Seller** wants to maximize the profits from trade: $U^S = p - c$.

Possible Outcomes

- All types of sellers are active $S_1 = \{H,M,L\}$
- Only a subset of sellers is active
  - $S_2 = \{M,L\}$
  - $S_3 = \{L\}$
  - $S_4 = \{H,M\}$
  - $S_5 = \{H,L\}$
- We want to check, which of these outcomes can be an equilibrium and what will be the resulting price if the seller makes a “take-it-or-leave-it-offer”.
- A “take-it-or-leave-it-offer” gives all bargaining power to the offering party.
- This could happen if there are many competing car buyers.

We check outcomes case by case.

Solving the Used Car Market Example

S_1: Could it be that all types of cars are offered?	• Maximum price the buyer will accept is $$E[v \mid S = S_1] = \frac{1}{3}\cdot 12 + \frac{1}{3}\cdot 6 + \frac{1}{3}\cdot 1 = \frac{19}{3}.$$ • This is strictly less than 9, i.e., what a good car dealer needs to break even -> thus, this cannot occur in equilibrium.
S_2: Could it be that only M & L are active?	• Maximum price the buyer will accept is $$E[v \mid S = S_2] = \frac{1}{2}\cdot 6 + \frac{1}{2}\cdot 1 = 3.5.$$ • This is strictly less than 4, i.e., what a medium car dealer needs to break even -> thus, this can also not occur in equilibrium.
S_3: Could it be that only L types are active?	• Maximum price the buyer will accept is $$E[v \mid S = S_3] = 1.$$ • This is enough for the L-types to break even -> thus, $S^*=S_3$ and $p^*=1$ is an equilibrium!

In Problem 3 of Problem Set 5 we will show that S_4 and S_5 can not be an equilibrium.

Adverse selection leads to inefficient market outcomes.

Interpretation

- Adverse selection occurs: “Bad types” drive good ones out of the market.
- Why? Because being pooled with worse types is unattractive, it lowers payoffs and makes participation less attractive.
- This can lead to full or (see problem set) partial market break-down.

Welfare Implications

- The market outcome is inefficient: Sellers realize only a surplus of 1 each, while it could be 2 with medium type and 3 with high type.
- Thus, there are gains from trade left unexploited; put differently, there are still Pareto-improvements possible.
- Why is a “competitive”[1] market not efficient here?
  -> The market is incomplete: there must be different prices for the different types for 1. Welfare theorem to apply; here, only one price for three different products.

1) The market we looked at is not competitive since the seller has market power. However, also a competitive market would not be efficient here, since sellers can never recoup their expenditures.

Further examples of adverse selection

Insurance markets

- People have better information about their health than an insurance company.
- Likely that unhealthy people seek to take out insurance.
- Premium increases.
- More healthy people cancel the insurance, and so forth.
- **Adverse Selection:** only unhealthy remain insured at a high premium.
- **Solution:** spreading/ pooling of risks. Examples: medicare in the US or compulsory insurance.

Credit markets

- Borrowers have better information about their default risk than banks.
- Likely that borrowers with a high default risk seek to take out a loan.
- Interest rate increases.
- Fewer creditworthy borrowers take out a loan, and so forth.
- **Adverse Selection:** only borrowers with high default risk take out a loan at a high interest rate.
- **Solution:** banks can share and use computerized credit histories ➔ privacy concerns!

We can (partially) solve the adverse selection problem and achieve a more efficient outcome

Possible solutions to the used car market problem

- Verification by independent experts: Both parties refer to an assessment of an independent institution which determines the quality of the cars.
  Example: Technical Control Board
- Self-selection mechanism: Uninformed party offers several contracts, and different types of sellers select different contracts (screening of sellers)
  Examples: Co-insurance or deductibles in insurance contracts
- Signaling:
  - The H-type car dealer gives a guarantee in form of a warranty. The dealer "signals" that he sells a high-quality car.
  - Such a warranty would be a promise to pay the buyer an agreed amount if the car turned out to be a middle- or low-quality car.
  - Sellers of high-quality cars can afford to offer such a warranty while sellers of low-quality cars cannot afford this.
    ➔ Sellers of good cars can distinguish themselves from sellers of bad cars!

1	What is asymmetric information?
2	Hidden Information: Adverse Selection
3	Hidden Information: Signaling
4	Hidden Action: Moral Hazard

Example of Signaling in the Labor Market.

Signaling through Education

A classic example of signaling is education.
The following is based on Spence (1973) and he was awarded the Nobel prize for the insights.

Two worker types

There are two types of workers: A high-ability type whose productivity is 10, and a low-ability type whose productivity is 5. The probability for each type is 0.5.

Edu-cation

Workers can invest in their education.
Assumption: It is less costly for high-ability workers to invest in education.
One unit of education, e, costs 1 for high-ability workers but 2 for low ability workers.

Wage payments without Signaling

Benchmarks

Firms Firms pay each worker according to their productivity.

This set-up implies that education does not increase the productivity of the worker. (Does not hold for Frankfurt School!)

Benchmark 1:
If the firm could observe the type of the worker, it would pay a wage of 10 to the high-ability worker and 5 to the low-ability one.

Benchmark 2:
If the firm could not observe the type of the worker and education was not possible, the firm would pay an average wage of 0.5(10+5)=7.5.

Separation Equilibrium (1)

Separation through education

- **Question:**
  Is it possible that workers can separate via choosing different levels of education?

In other words, is there a constellation so that the following holds:

1. All low-ability workers choose an education level e=0.
2. All high-ability workers choose an education level e*>0.
3. Firms pay a wage of 10 for a worker with e=e* and 5 for a worker with e=0.

For this to be an equilibrium, no agent should have a profitable deviation incentive.

Separation Equilibrium (2).

Deviation incentives

- High-ability worker:

  He prefers e=e* to e=0 if and only if

$$10 - 1 \cdot e^* > 5 - 1 \cdot 0$$

  or e* < 5.

- Low-ability worker:

  He prefers e=0 to e=e* if and only if

$$5 - 2 \cdot 0 > 10 - 2 \cdot e^*$$

  or e* > 2.5.

Consequence:

For e* between 2.5 and 5, there is a separating equilibrium, in which the high-ability worker can signal his type through education.

Separation because ability is negatively correlated with education costs.

Remarks

- Intuition:

  The high-ability worker chooses such a high level of education that the low-ability worker has no incentive to mimic due to her higher costs for education.

- Signaling leads to a reduction in welfare! The (average!) wage is unchanged but education is costly. Therefore, costs are higher but the productivity of the workers is the same.

  The separating equilibrium which is least costly is the one with e^* close to 2.5.

- Paradoxically, the result arises although education does not improve the ability of the worker. However, as the costs for education are correlated with the worker's ability, signaling works.

We specify a situation similar to our former labor contract example.

Example: Sales Agent

Situation

- A sales agent is responsible for a certain region for her company.
- How much sales she can make depends on how hard she works but also on many other factors she cannot influence and which cannot be observed by her or her company.
- The company wants to induce her to work hard but cannot (at reasonable cost) monitor her effort level.
- The sales agent has no other source of income. Thus, she would like to see some stability in her income stream.
- The company, having many sales agents serving many different regions is less concerned about the variance of sales in any particular region, since, due to the law of large numbers, it can benefit from portfolio effects.
- What is the optimal contractual arrangement between the company (the "principal") and its (sales) agent?

Again, we try to gain insights from a parameterized model.

Example: Sales Agent, cont.

- The sales agent can make no sales, medium sales, and high sales. The (gross) profits for the principal are $\pi_L = 0$, $\pi_M = 100$, and $\pi_H = 400$.
- The probability of each case depends on the agent's effort *e*, which can either be high or low:

Agents effort	$prob_L = prob(\pi = \pi_L)$	$prob_M = prob\ (\pi = \pi_M)$	$prob_H = prob\ (\pi = \pi_H)$	$E[\pi \mid e]$
high	0.1	0.3	0.6	270
low	0.6	0.3	0.1	70

- The principal cares only about expected profits, net of the agent's salary w.
- The agent's effort cost is 5 when working hard and 0 when not working hard.
- The agent gets utility from her salary and disutility from working hard:

$$U_A(w,e) = \sqrt{w} - e, \text{where } e = 0 \text{ or } e = 5$$

- This implies that the agent is risk averse.
- Finally, the agent has a "reservation utility" of 9, i.e., if the utility resulting from the principal's offer is below 9, the agent will not participate (i.e., quit).

The challenge is to make the agent work hard by choosing the right contract offer.

The Sales Agent Problem

- With symmetric information, the agent should work hard.
- Why? When working hard, the agent produces an expected surplus of 270 for the principal.
- The compensation payment demanded by the agent is:

$$\sqrt{w} - 5 \geq 9 \leftrightarrow w \geq (9+5)^2 = 196 < 270.$$

- Instead, letting the agent shirk implies that the principal obtains only 70 but needs to pay a wage so that $\sqrt{w} \geq 9$, which leads to w=81. The principal would make a loss.
- Optimal contract if the principal can observe the effort level: {e=5,w=196}
- However, with **asymmetric information** the principal cannot enforce the agent to work hard. If he would just offer w = 196, a rational agent just takes the money and shirks (i.e., chooses e = 0).
  ➔ Then, the principal makes a loss of 126 (=70 – 196).
- Letting the agent shirk does not work either as this also results in a loss, as derived above.

We can achieve the first best with a risk neutral agent by selling the project to the agent.

Benchmark Case: The Agent is Risk Neutral

- Assume on this slide that the agent is risk neutral, i.e., $U_A = w - e$, and that the high effort level is $e = 25$ (instead of $e = 5$) and the reservation utility is *81*.
- In this case, there is a simple solution to the problem: The principal offers the agent the following contract: "You (the agent) pay me some amount upfront. In return, you get all the profits you are able to realize in the market."
- This is like a "franchise contract" and obviously sets very strong incentives for the agent since, if the agent agrees, she can keep 100% of everything she earns.
- The agent is "residual claimant" of all returns and since the expected returns exceed the cost, she will choose the efficient effort level (e = 25 here).
- What is the upfront payment X? Reservation utility = 81 ≤ 270 – X – 25.
- With X=164, the agent accepts.
- Selling the project to the agent implies that the agent becomes her own principal, solving the asymmetric information problem.

There is a fundamental trade-off between incentives and insurance.

Trade off between incentives and insurance

- A fundamental problem in economic analysis as well as in abundant real world situations is the trade-off between insurance and incentives.

An optimal contract should:

Insure the risk averse agent (the principal should bear all the risk, since he is willing to do so).

This calls for “flat” contracts, e.g. fixed price contract (same payoff for all realized sales levels = full insurance for the agent).

Provide incentives for the agent to exert effort.

This calls for “steep” contracts (high payments only if the realized sales levels are high = selling the project to the agent).

Both targets contradict! They cannot be fully satisfied at the same time!

An optimal contract must satisfy the constraints of voluntary participation and incentive compatibility.

How could the optimal contract look like?

- It cannot be a fixed payment to the agent.
- To induce the agent to work hard, the agent has to be compensated to accept risky payoffs, which is costly for the principal.

The principal has to take care of **two restrictions** when proposing a contract:

(1) Participation of the agent: If he offers too little (in terms of expected utility), the agent will not accept the contract (the agent can not be forced to participate – she could always quit and realize her reservation utility).

(2) Incentive compatibility: The payoffs, conditional on the verifiable outcomes, must be structured such that the agent indeed prefers to work hard, instead of shirking.

Formally, finding the optimal contract is a constrained maximization problem.

Finding the optimal contract formally

Con-straints

- *(PC)* : given the salary scheme, if the agent chooses high effort, she must get at least her reservation utility of 9 ➔ it must be profitable for the agent to accept the contract!

- *(IC)* : given the salary scheme, if the agent works hard, she must get at least as much as if she shirked ➔ it must be more profitable for the agent to undertake $e=5$ instead of $e=0$!

Optimal contract

- The action "low effort" $(e=0)$ can never be optimal as it yields an expected payoff of 70 but the wage to the agent must be at least 81
- ➔ Consequence: high effort must be optimal
- What is the minimum salary scheme w_i ($i = H,M,L$) the principal must offer the agent to induce the agent to spend high effort ($e=5$)?

We set up a Lagrange function to find the optimal contract in our example.

Solving for the optimal contract

- The principal needs to induce *e = 5* to realize an expected payoff of *270.*
- Thus, we can reformulate the problem as a minimization problem: what is the minimum the principal has to pay the agent to make her work hard?

(Obj. Funct.) $$\frac{1}{10}w_L + \frac{3}{10}w_M + \frac{6}{10}w_H \rightarrow \min_{w_L, w_M, w_H}$$ subject to:

Optimization Problem

(PC) $$\frac{1}{10}\sqrt{w_L} + \frac{3}{10}\sqrt{w_M} + \frac{6}{10}\sqrt{w_H} \underbrace{-5}_{\text{effort}} \geq \underbrace{9}_{\text{reservation utility}}$$

(IC) $$\underbrace{\frac{1}{10}\sqrt{w_L} + \frac{3}{10}\sqrt{w_M} + \frac{6}{10}\sqrt{w_H} - 5}_{\text{utility from } e=5} \geq \underbrace{\frac{6}{10}\sqrt{w_L} + \frac{3}{10}\sqrt{w_M} + \frac{1}{10}\sqrt{w_H}}_{\text{utility from } e=0}$$

The participation constraint and the incentive constraint are binding in the optimum.

Binding Constraints

- Why is the participation constraint binding?
  - *Suppose not: Then the principal could lower the expected payoff for the case that the agent takes the "correct" action. The agent would still participate and the principal has saved money!*

- Why is the incentive constraint binding?
  - *Suppose not: This would mean that the agent earns strictly more when choosing the "correct" action compared to choosing the "wrong" action. Then, the principal could (slightly) lower the payoffs for choosing the "correct" action and still induce the correct action – and again save money!*

We set up a Lagrange function to find the optimal contract in our example.

Solving for the optimal contract

- To find a solution to the principal's objective function subject to the participation (PC) and incentive compatibility constraint (IC) we need to set up the Lagrangian.

Lagrange Function

$$L = \frac{1}{10} w_L + \frac{3}{10} w_M + \frac{6}{10} w_H$$

$$+ \lambda \left(\frac{1}{10}\sqrt{w_L} + \frac{3}{10}\sqrt{w_M} + \frac{6}{10}\sqrt{w_H} - 14 \right) + \eta \left(-\frac{5}{10}\sqrt{w_L} + \frac{5}{10}\sqrt{w_H} - 5 \right) \rightarrow \min_{w_L, w_M, w_H}$$

We can find the solution by analyzing the first order conditions.

Solution

- The solution involves maximizing the Lagrangian function with respect to w_L, w_M and w_H.
- The details on how to solve for the optimal contract are provided on a separate file, available on the course webpage:
- The optimum contract is:

Deriving the optimal contract

"The agent gets w_L = *29.469 in case of zero sales*, w_M = *196 in case of sales of 100, and* w_H = *238.040 in case of sales of 400.*

Agent

- Accepts the contract and chooses high effort, $e = 5$.
- Receives (just) her reservation utility, $U_A^R = 9$.

Principal

- Has to pay expected salaries of *204.571.*
- Implies expected profits of *65.429.*

Welfare loss.

Discussion

- **Welfare loss due to asymmetric information**: There is a welfare loss, since the agent has to bear some risk ($w_L = 29.469$, $w_M = 196$, $w_H = 238.040$). Why?
- **Compromise between incentives and insurance**: The risk taken by the agent is lower than in a full powered incentive contract where she would be sold the whole project (after buying the project, the agent's salary scheme would be $w'_L = 0$, $w'_M = 100$, $w'_H = 400$).
- **What is the first best? Complete information**: the principal induces the agent to spend high effort and the agent just accepts the contract ➔ The principal gets $E[\pi|e] = 270$ and the agent accepts if the salary satisfies $\sqrt{w} - 5 \geq 9$ ➔ $w = 196$.

Welfare loss.

Discussion

- In this case, the welfare loss is reflected in **lower profits for the principal**: in the **first best**, the principal needs to pay only *196,* implying an expected profit of *74.*

➢ With **asymmetric information**, the expected salary that the principal has to offer the agent to induce her to spend high effort is $0.1*29.469 + 0.3*196 + 0.6*238.04 = 204.571$ which yields an expected payoff of *65.429.*

➢ The welfare loss is thus *74 – 65.429 = 8.571.*

Conclusion of the principal-agent model.

Conclusions

- Hidden action leads to "**moral hazard**": costly efforts will not be undertaken if they can not be sanctioned.
- Thus, the principal (the one who proposes the contract) must set some **incentive for the agent to undertake the effort**.
- Setting **incentives is usually costly**.
- A very frequent form of costs are **limitations** to the degree of **insurance** the agent can get.
- Thus, there is a basic **trade-off** between (efficient) insurance and incentives.

Exam.

- Solving a Lagrangian with 3 unknowns is not relevant for the exam!
- **BUT**, you should be able to solve the principal-agent problem with two unknowns. Example: Problem 4 of Problem Set 5.
- There, the participation and the incentive compatibility constraint are already a system of two equations with two unknowns; (we know that in the optimum the conditions are binding, i.e., they must hold with equality).

Key Points

Summary

1. Asymmetric information means that one player has relevant information while the other has not. “Relevant information” means that an efficient contract would condition on this information.
2. Generally, there are two types of asymmetric information:
   - before contract = hidden information, leads to adverse selection
     example: car dealers -> good products might be driven out of the market; can be (partially) solved through signaling
   - after contract = hidden action, leads to moral hazard
     example: sales agent -> incentives required, which destroy optimal risk sharing

Data Collection

Topics

- Data, sampling
- Correlation vs causality
- Empirical research design
- Statistical evaluation

Structure of these slides

1. Introduction to "econometrics": using data to test theories
2. Potential outcome model: formal concept of causality
3. Randomized experiments
4. Types of data
5. Example: making use of data

Econometrics

Data analysis and econometrics

As social scientists (in business, economics, psychology, management, and finance) we are interested in how society works.

Observing the world is the basis for all social science research.

- Theories explain observed phenomena.
- Theories give rise to testable hypotheses.
- Without theory, what should we test?
- But without empirics, what is the relevance of theories?

Exchange between **theory and empirics** is the heart of the scientific process.

Data analysis and econometrics

Many decisions in economics and social sciences overall require understanding the relationships between variables.

Decisions often require **quantitative answers** to **quantitative questions**.

- To be able able to give policy advice.
- To do cost-benefit analyses.
- ...

Theory (across social sciences) suggests important relationships, often with policy implications, but virtually never suggests **quantitative magnitudes** of **causal effects**.

This is where **Econometrics** comes in:
Econometrics is the science and art of using (economic/social science) theory and statistical techniques to analyse data.

What is the goal?

Suppose we are interested in the connection between an outcome variable y (e.g. workers effort, job satisfaction, ..) and a variable x which may affect y (e.g. wage, the size of bonus payments, whether the firm uses performance pay or not)

Let e be a variable which describes all other determinants of y that we do not observe. Then we can denote the relationship between y and x as

$$y = f(x, e)$$

Key aim: Understand this function and learn about it by analyzing data

Data analysis and econometrics

Learning about f means: **establishing** ("identifying") and **quantifying causal relationships** between the explanatory variable x and the outcome variable y.

Many examples!

- Does another year of education change earnings? How much?
- Do cigarette taxes reduce smoking? How much?
- Does a lower interest rate stimulate the economy? How much?
- Does development aid reduce poverty? How much?
- Does social distancing reduce Covid-19 infections? How much?
- Does online teaching improve study results? How much?
- Does more police reduce crime? How much?
- ...

Why Data?

- Data is objective
- Oftentimes we miss dependencies without data
- Anything else is guessing and may lead to no new insights
- Subjective reasoning is oftentimes the same as using very small datasets
- May allow to provide quantitative answers to questions

Why not data?

- might not be available
- takes time /is expensive
- Might be misleading if wrongly interpreted

Data analysis and econometrics

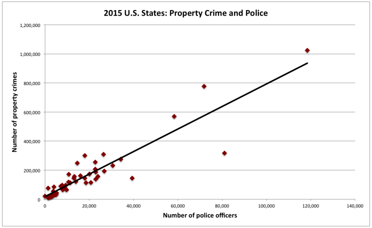
Although 'common sense' may suggest a direction - or a *qualitative* answer - for these relationships, it cannot provide **quantitative** answers.

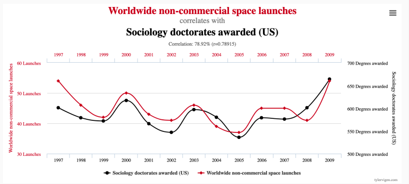
- The quantitative importance of the relationship must be learned empirically - that is, by analysing **data**.
- Because we use data to answer quantitative questions, our answers always have some **uncertainty**: A different set of data would produce a different numerical answer.
- Our conceptual framework for the analysis - that is, our empirical methods - therefore need to provide both a **numerical answer** as well as a measure of how precise the answer is (**inference**).

Correlation and causality

Data may show positive correlations, for example:

- Level of education and health status at old age.
- Smoking behaviour during pregnancy and baby's birthweight.
- Size of the police force and crime.
- Non-commercial space launches and number of sociology doctorates (US).





What do we learn from these correlations?

Correlation and causality

Causality means that a specific action/variable leads to a specific consequence.

- Example: Drinking water makes one less thirsty.

Causal effects are very different from correlational relationships!

- Example: Bringing your umbrella in the morning is correlated with a higher probability of rain in the afternoon.
- Further examples? Let's discuss!

Correlation and causality

In order to understand how society works in general and to make evidence-based policy recommendations, isolating (or: **identifying**) **causal effects** from correlations is crucial.

- Causality is difficult to establish (correlation is easy).

By using different types of data, and more or less "fancy" statistical methods, this is what **Econometrics** does!

- Study relationships between variables.
- Evaluate policies.
- Forecasting.

Similar methods are used **across disciplines** within and outside the social sciences:

- Economics, Sociology, Management, Political Science.
- Epidemiology, Psychology, Medical Science, etc.

The causality challenge

We need to understand causalities to answer questions such as:

1. What *would have* happened (ex-post evaluation)...
2. What *would* happen (ex-ante prediction)...
    - ... if someone's diet had been/would be healthier?
    - ... if the interest rate had been/would be higher?
    - ... if a 'lockdown' had not been implemented?
    - ... if one had not/would not attend university?

Answering such questions requires learning about unobserved outcomes, because we only observe *one* **potential outcome**.

Ex.: Do people earn more if they complete an MBA at university?

- We observe income for people given their education, not what their income would have been if they had a different education.

$\rightarrow$ This is what we call the **unobserved counterfactual**.

The causality challenge

One instinctive approach might be to compare income of individuals **with** or **without** an MBA. Discuss: Good approach?

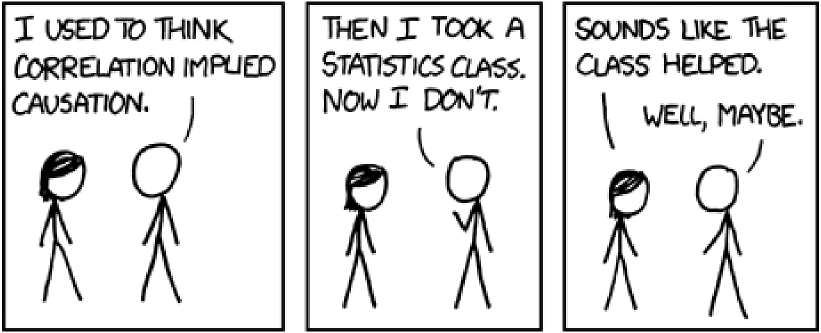
Probably not! Correlation between X and Y can be caused by:

1. A causal **effect of X on Y**. Obtaining an MBA causally impacts an individual's wages.
2. A causal **effect of Y on X**. Individual's with a higher wage can afford to pursue an MBA.
3. Omitted variables: **Z affects both X and Y**. Individuals with a higher academic ability may earn more and be more likely to have an MBA.

Distinguishing between the three is difficult, but crucial e.g. for **evidence-based decisions** (in policy or otherwise).

$\rightarrow$ Econometrics offers tools to achieve that!


Correlation
Causality


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Potential outcome framework

Potential outcomes framework

- Let $c_i \in \{0, 1\}$ be a dummy variable indicating whether the person i is treated or not
- What we would like to know is: what is the value of y_i
  if $c_i = 1$ (person i is treated)
  if $c_i = 0$ (person i is not treated)
- Let this potential outcome be

$$y_i(c_i) = \begin{cases} y_{i1} & \text{if } c_i = 1 \\ y_{i0} & \text{if } c_i = 0 \end{cases}$$

- The causal effect of c_i on y_i is given by $y_{i1} - y_{i0}$ for person i

Example

- Health outcome as a function of health insurance:

	Kamal	Amy
Potential outcome without insurance: y_{i0}	3	5
Potential outcome with insurance: y_{i1}	4	6
Treatment (insurance status chosen): c_i	1	0
Actual health outcome: y_i	4	5
Treatment effect: $y_{i1} - y_{i0}$	1	1

Causal inference as a missing data problem

- Problem: We only observe one potential outcome per person
- "Counterfactual" is missing ...
- And: c_i is not randomly assigned, but chosen by i
- In example:
  - Amy has better health more generally and does not choose to insure herself (which may be reasonable)
  - Kamal is worse off health-wise and may opt for insurance

Naïve valuation of the treatment "insurance"

- Comparison of observed insured vs not:

  $y_{Amy} = y_{Amy0}$

  $y_{Kamal} = y_{Kamal1}$

- So:

$$y_{Kamal} - y_{Amy} = -1$$

- Observed difference suggest health insurance is counterproductive

Selection bias

- $y_{Kamal} - y_{Amy} = y_{Kamal1} - y_{Amy0}$

$$= \underbrace{y_{Kamal1} - y_{Kamal0}}_{+1} + \underbrace{y_{Kamal0} - y_{Amy0}}_{-2}$$

- First term is true causal effect for Kamal
- Second term is baseline difference in frailty between Kamal and Amy
  - Baseline difference biases naïve valuation!

Selection bias more generally

- Naïve comparison: Δ

$$\begin{aligned}
\Delta &= E[y_{i1}|c_i = 1] - E[y_{i0}|c_i = 0] \\
&= E[y_{i1}|c_i = 1] - E[y_{i0}|c_i = 0] + E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 1] \\
&= E[y_{i1}|c_i = 1] - E[y_{i0}|c_i = 1] + E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 0] \\
&= E[y_{i1} - y_{i0}|c_i = 1] + E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 0] \\
&= \underbrace{E[y_{i1} - y_{i0}]}_{\text{Causal Effect}} + \underbrace{E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 0]}_{\text{Selection Bias}}
\end{aligned}$$

$$\Delta = \underbrace{E[y_{i1} - y_{i0}]}_{\text{Causal Effect}} + \underbrace{E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 0]}_{\text{Selection Bias}}$$

- Selection effect: If the potential health outcome of those with health insurance differed from the potential health outcome of those without health insurance, the estimate of the treatment effect is biased.

- Here: Plausible if e.g. less healthy individuals more likely to sign up for health insurance

- Another example: Job training program effect on wages

$$\Delta = \underbrace{E[y_{i1} - y_{i0}]}_{\text{Causal Effect}} + \underbrace{E[y_{i0}|c_i = 1] - E[y_{i0}|c_i = 0]}_{\text{Selection Bias}}$$

- Unless we know how health insurance was assigned, cannot make any further inference
- Essentially all of causal inference is about correcting or eliminating selection bias

Randomized Experiments

Randomised experiments

To start understanding how the **self-selection** problem can be solved, let us consider randomised experiments.

- A **treatment** (e.g., a drug in medical trials) is assigned **randomly** across individuals, often using a lottery.
- The outcome for this group (e.g., a measure of health) is compared to the randomly selected **control group**.
- The magic of randomisation: **Eliminates the selection problem** - the possibility of systematic differences between those who receive a treatment and those who do not.
- The only **systematic reason** for differences in the outcomes between treatment and control groups is the **treatment** itself and we can isolate the **causal effect** of the treatment.

Randomised experiments

We can use the **potential outcomes framework** to better understand how experiments isolate causal effects of the treatment:

- Treatment D is **assigned randomly** across individuals.
- So, treatment assignment is statistically **independent of potential outcomes**

$$(Y_{0i}^*, Y_{1i}^*) \perp D_i$$

- Random assignment solves the **self-selection problem**:

$$\mathrm{E}[Y_1^*] = \mathrm{E}[Y_1^* | D = 1] = \mathrm{E}[Y_1^* | D = 0]$$
$$\mathrm{E}[Y_0^*] = \mathrm{E}[Y_0^* | D = 1] = \mathrm{E}[Y_0^* | D = 0]$$

- This implies $ATE = ATET$. Intuition: The treated are a random subsample of the population.

Randomised experiments

We can analyse data from experiments by comparing **differences in means** (the difference-in-means estimator). With randomised treatment, we are interested in $\mathrm{E}[Y_1^*|D=1] - \mathrm{E}[Y_0^*|D=0]$.

For treatment status D_i (=0 or =1), the **observed outcome** is:

$$Y_i = (1 - D_i)Y_{0i}^* + D_i Y_{1i}^*$$

We can **estimate** $\mathrm{E}[Y_1^*|D=1]$ and $\mathrm{E}[Y_0^*|D=0]$ by using **sample means**:

$$\hat{\mathrm{E}}[Y_1^*|D=1] = \frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i} \text{ and } \hat{\mathrm{E}}[Y_0^*|D=0] = \frac{\sum_{i=1}^n (1-D_i) Y_i}{\sum_{i=1}^n 1 - D_i}$$

The **resulting estimator** for the treatment effect(s) is

$$\widehat{ATE} = \widehat{ATET} = \frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n (1-D_i) Y_i}{\sum_{i=1}^n 1 - D_i}$$

Randomised experiments

The concept of an ideal randomised controlled experiment is useful because it gives a clear definition of a causal effect.

$\rightarrow$ *If experiments allow us to isolate causal effects, why do we not always conduct randomised controlled experiments?*

- In practice, often not possible.
- Financial reasons, practical reasons, ethical reasons.

Thus, we need to rely on (other) statistical methods - but ideal experiments are still the benchmark when thinking about how to isolate causal effects.

Types of Data

Randomised experiments: Examples

Do smaller class sizes improve student learning?

- STAR class size experiment in Tenessee (1990s).
- Pupils in primary school were **randomised** into small classes, large classes and classes with a teaching assistant.
- Suffered from **partial compliance**: Principals switched pupils between classes & some evidence that the principals assigned the best teachers to the large classes.

Randomised experiments: Examples

Is there discrimination in the labor market?

- *Audit studies* send out resumes with randomised contents.
- Bertrand and Mullainathan (2004): Send pairs of resumes with same characteristics but either a "white-sounding" name (Emily, Greg) or a "black-sounding" name (Lakisha, Jamal) - constructed from official records.
- Count number of times each fictitious applicant receives a call-back for job interview. Result: "White" names receive 50% more callbacks. Strong evidence of discrimination!

Many (!!) follow-up studies: Ethical issues of experimental design become more and more important.

Non-experimental methods

Often, experiments are not **feasible** (unethical, impractical, expensive, etc.) and alternative approaches are necessary:

- **Natural experiments:** Random variation that occurs without the researcher's intervention (e.g. medical school admission, military draft, Germany's reunification).
- **Quasi-experiments:** Situations that yield quasi-random variation, for example due to policy changes (comparison of groups, regions, states, countries over time) or cut-off rules (test-scores, birth-dates, etc.).

Sources of data

Experimental data:

- Data that are generated by randomised controlled experiments.
- Causal effects obtained by random assignment of treatment in a controlled experimental environment.

Observational data:

- Data that are generated without a controlled experimental setting - "real-world" data.
- Causal effects obtained by other sources and with the help of **econometric tools** outside of controlled experimental settings.
- Types of data: **cross-sectional** data, **time series** data, **panel** (or longitudinal) data.

Which type of data do you need?

- This depends highly on what question you try to answer and what constraints you are facing
  - Getting data may be expensive, either it may require you to buy observational data (there is some publicly available data though!) or it may require you to measure something yourself which often is not resource neutral
- In answering this question, we will discuss internal validity and external validity as organizing principles
  - Internal validity: confidence in how much we trust that the empirical relationship that is studied is correctly interpreted in its context
  - External validity: confidence in how much we trust that the empirical relationship that is studied applies more generally
  - Try to maximize both, but often there is conflict between them

Example 1

- You would like to know if a certain training program improves the productivity of a particular group of employees
  - You may focus more on internal validity if you are not concerned about the effect of the training on productivity outside the targeted group of employees
  - Maybe you will try to do a RandomizedControlTrial and measure productivity yourself directly

Example 2

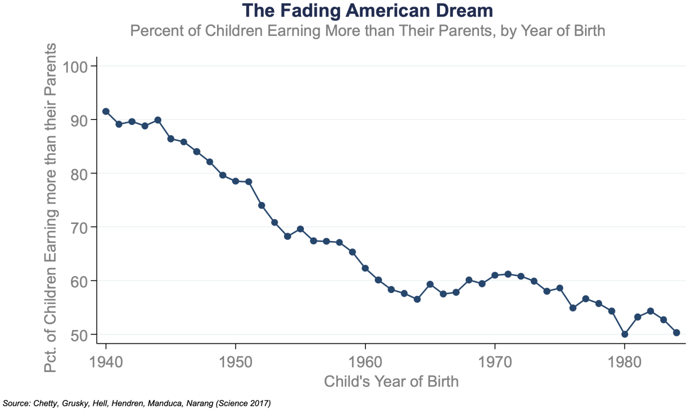
- You would like to know if increasing interest rates hurts economic demand
  - You may focus more on external validity because you are interested in a general phenomenon that should apply everywhere
  - External validity: confidence in how much we trust that the empirical relationship that is studied applies more generally
  - You may want to use observational data that allows you to show how economic demand is affected in various economic sectors

Data collection and data analysis

- Another important factor for data collection is data analysis
- If data is supposed to help you answer a question,
  then you should know how you would like to analyze the data
- This may have direct consequences for data collection
- Therefore, we will also recap regression analysis in the next set of slides

Example

Example: intergenerational mobility



Data: time series of earnings data from public records

Economics and its main problem

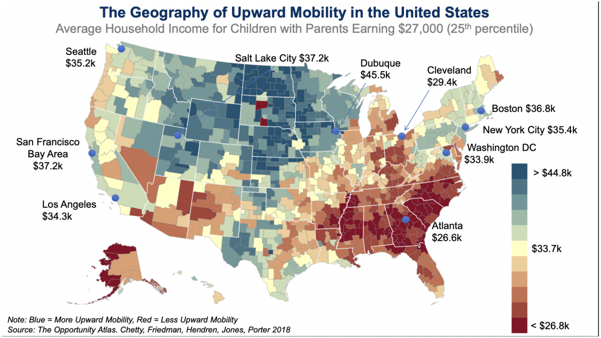
- Central policy question: why is the American dream fading?
- Difficult! Many many changes occurred over time, yet we have little data
- Economists *used* to develop mathematical models to help explain phenomenon and make policy recommendations mere based on theory
- Problem: untested theories cause persistent scientific disagreement

  which leads to politicization of questions that have scientific answers
- Need for empirical research to have a better understanding!
  - This is great motivation for a research paper or, actually, research agenda!

What determines upwards mobility?

- Start with what theory predicts and/or further “empirical” facts.

  Here, the latter:



- Understand causes by understanding geographic variation?

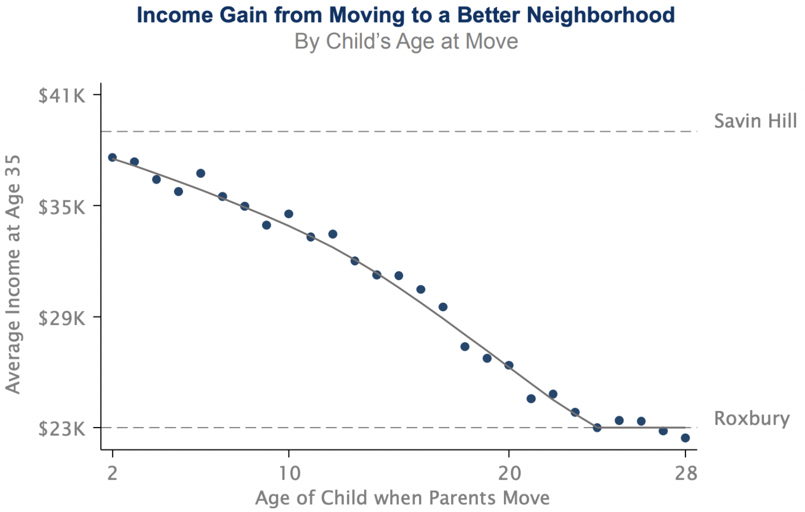
Neighborhood vs sorting

- Two potential explanations for geographic variation:
  1. Different people live in different places
  2. Places have a direct effect on upwards mobility

- Assume we are mostly interested in the direct effect of the place

  How would you ideally study it?

Approximation of experiment: moving and age of children



Key assumption behind direct place effect

- Timing of move is unrelated to other determinants of child's outcomes
- May not be true!
  - Parents who move to good places when kids are young may be different
- If assumption does not hold: age effect could "merely" be about sorting!
- Additional evidence may increase our confidence in assumption (or not)
- Issue: correlation vs causality
  - Age-of-kids and upward mobility may be correlated but this does does not have to imply that places causalty effect upward mobility. Sorting may be true causal effect and may be correlated with age-of-kids.

Correlation does not imply causation!

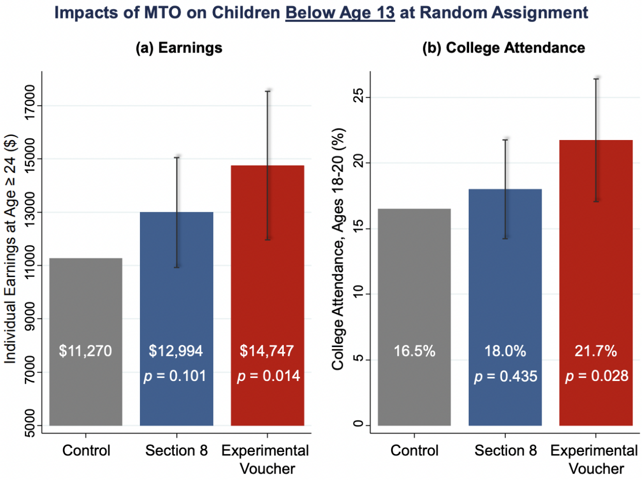
- Mere correlation between age-of-kids and upward mobility:
  - May not be a causal relationship because of omitted variables
  - Here, for instance, sorting.
  - Under what conditions would a causal interpretation be valid?
  - Random assignment of “treatment”
  - Treatment: place you live upwards mobility question

Moving to Opportunity Experiment

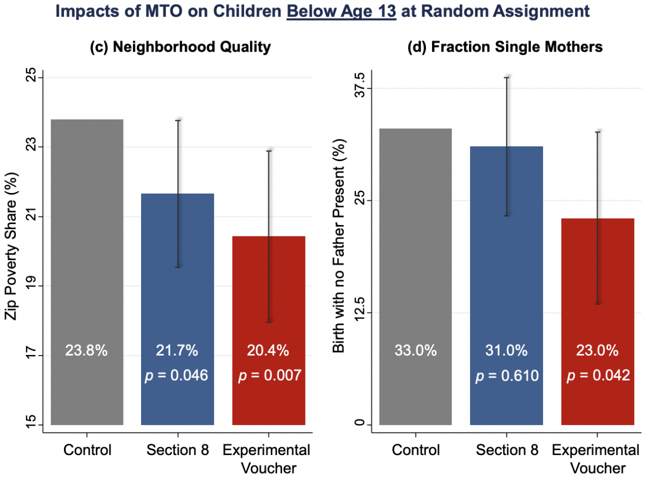
- 1994-1998 in Baltimore, Boston, Chicago, LA, New York
- 4,600 families were randomly assigned to one of three groups:
  - Experimental: offered housing vouchers restricted to low-poverty (¡10%) Census tracts
  - Section 8: offered conventional housing vouchers, no restrictions
  - Control: not offered a voucher, stayed in public housing
- Compliance rate: 48% of experimental group used voucher, 66% of Section 8
- Does MTO improve upwards mobility?

  If yes, causal evidence for direct place effect!

MTO does affect upwards mobility 1/



MTO does affect upwards mobility 2/



Data Collection

Structure of these slides

1. Univariate Regression
2. Parameter Interpretation
3. OLS Assumption
4. Multivariate Regression

Regressions

Recall: the goal

Suppose we are interested in the connection between an outcome variable y (e.g. workers effort, job satisfaction, ..) and a variable x which may affect y (e.g. wage, the size of bonus payments, whether the firm uses performance pay or not)

Let e be a variable which describes all other determinants of y that we do not observe. Then we can denote the relationship between y and x as

$$y = f(x, e)$$

Key aim: Understand this function and learn about it by analyzing data

How do we estimate relationships between variables?

We use regressions:

- They provide useful approximations to conditional expectation functions
- And conditional expectation functions are a powerful tool to predict outcomes

Regressions are first order important to understand conditional expectations and to derive predictions.

Conditional Expectation Function

- We are interested in the conditional expectation function (CEF) of y given x.
- For example performance given that the firm pays a bonus for managers.
- This is: $E[y|x]$
- Interpretation: $E[y|x]$ can be interpreted as the mean of y among all observations that share the same value(s) of x

Conditional Expectations and Regressions

- The CEF gives the population average of y for the group of people having the same x
- If I know the CEF, I can make predictions which value y would take for different values of x
- Regressions and several machine learning algorithms are tools to approximate the CEF
- Typically, we will not know the functional form of the CEF. But we can try to approximate it

Linear Regressions

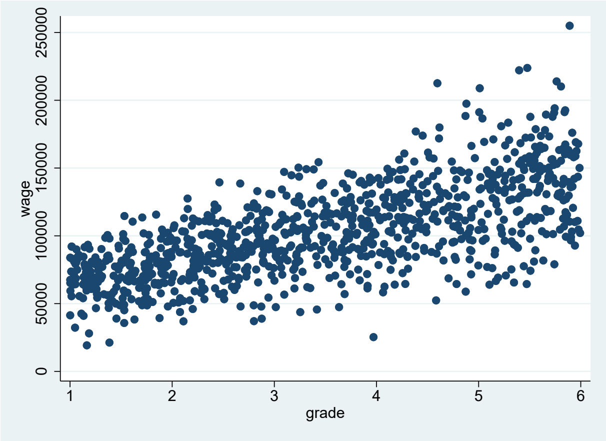
- Suppose there is an outcome Y with observed realizations $(y_1, \cdots, y_N)$
- And an independent variable (predictor) X with observed realizations $(x_1, \cdots, x_N)$
- Linear (univariate) regressions assume that

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- i is the unit of observation
- β_0 is the intercept
- β_1 is the slope
- ϵ_i is the error term
- These three variables are unknown and we aim to estimate them.

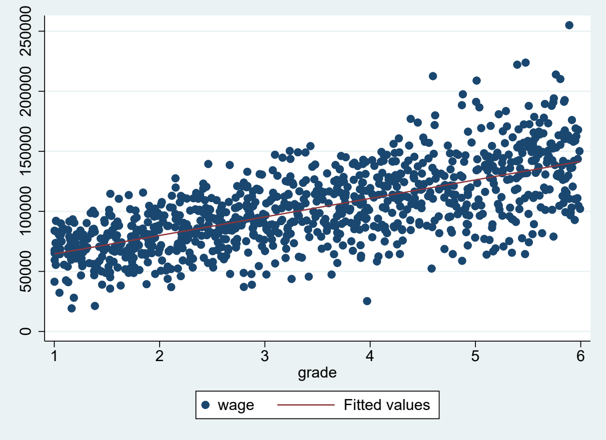
Linear Regressions: Example

Y First wage, X grade in my class: Scatterplot, 1000 Observations



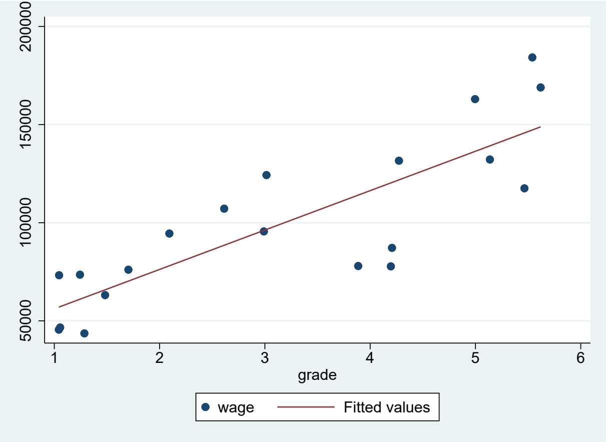
Linear Regressions: Example

What would a regression do?



Linear Regressions: residuals

Recall: The ordinary least squares (OLS) estimates β_0 (intercept) and β_1 (slope) in order to minimize (squared) residuals



Parameter Interpretation

Interpretation of $\hat{\beta}_1$

- $\hat{\beta}_1$ estimates the linear relationship between x and y
- This relationship has 4 important dimensions:

1. Sign: when x increases by one unit, does y increase or decrease?
2. Magnitude: when x increases by one unit, how much does y change?
3. Statistical significance: given the uncertainty of estimating β_1, is the estimated value of β_1 statistically meaningful?
4. Correlation vs causation: does $\hat{\beta}_1$ capture the causal effect of x on y or merely reflects a correlation between the two variables?

Marginal effects: Interpretation 1 and 2

Assume the model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- The coefficient $\hat{\beta}_1$ estimates the **partial** or **marginal effect** $\frac{\partial y}{\partial x}$: *How much does y change (in expectation) if x changes by one unit?*

  Remember that $\hat{\beta}_1$ estimates the slope of the regression line!

- Example: How much do earnings change if with one more year of education?
- Important: Whether or not this change is statistical significant and causal are other questions for now

Interpretation 3: Uncertainty and statistical significance

- $\hat{\beta}_1$ estimates the relationship between x and y
- But note: $\hat{\beta}_1$ is a random variable. Assume you sample data again, your estimate of β_1 may be somewhat different
- We need to take this uncertainty into account
- This is what we do in hypothesis testing
- Typically: Null hypothesis, H_0, is that $\beta_1 = 0$
- Alternative hypothesis, H_1, is that $\beta_1 \neq 0$
- Compute probability of obtaining the estimate ($\hat{\beta}_1^{OLS}$), *conditional on H_0 being true.*
- Reject the null hypothesis if this probability is small, with some pre-defined threshold (typically 5%).

Interpretation 3: Hypothesis testing & statistical significance

We call this probability the **p-value**: The probability of observing an estimate as or more extreme as the one we obtained if H_0 is true.

$$\text{p} - \text{value} = Pr\left[|\hat{\beta}_1 - \beta_1^0| > |\hat{\beta}_1^{OLS} - \beta_1^0|\right]$$

- β_1^0 is our chosen **value** for the null hypothesis (often 0).
- $\hat{\beta}_1^{OLS}$ is the **estimate** obtained from OLS.
- $\hat{\beta}_1$ is the OLS-**estimator**, which is a random variable.

Interpretation 4: Zero conditional mean assumption 1/

- $\hat{\beta}_1$ estimates the (linear) causal effect of x on y

  Is the estimator unbiased?

- Recall from statistics: $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \epsilon_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})}$
  - The observed relationship between x and y can be expressed as the true causal effect of x on y (which may be "$= 0$") plus the true covariance of ϵ and x (scaled by the inverse of the variance of x)

Interpretation 4: Zero conditional mean assumption 2/

- For $\hat{\beta}_1 = \beta_1$, it needs to be the case that $\sum_{i=1}^{n} \epsilon_i (x_i - \bar{x}) = 0$
  - $\hat{\beta}_1$ does not necessarily estimate unbiased causal effect of x on y
  - Only if zero condition mean assumption holds, i.e., $E(\epsilon|x) = 0$, which posits that error term and explanatory variable are independent of each other, does $\hat{\beta}_1$ estimate the true causal effect

    There are no omitted variables, there is no selection bias etc.
  - If $E(\epsilon|x) \neq 0$, $\hat{\beta}_1$ estimates the correlation between x and y that may be caused by third factors / omitted variables (that are in ϵ)

Example: Bertrand and Mullainathan (2004)

Design

- Create fictitious resumes of job applicants
  - make them realistic, representative (difficult, don't underestimate!)
  - for given occupations, create high and low quality resumes

Design

- Create identities of applicants
  - address, phone number, *names*
  - names are crucial to manipulate race
  - use uniquely black and uniquely white names (confirm in separate study)
  - applicants in each race/sex/city/resume quality cell are allocated the same phone number to be able to track callbacks.
  - fictitious addresses
- Identify open positions

Design

Key design idea:

- For a given job ad, randomly sample 4 resumes (2 high quality, 2 low quality)
- One of the high and one of the low quality applications is randomly assigned a black-sounding name
- Send out applications and measure call-back (phone-calls and emails)

⇒ Exogenous variation in race of applicant, everything else kept constant (not directly, though...)

⇒ Precise measurement of employer behavior

Results

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (*p*-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Notes: The table reports, for the entire sample and different subsamples of sent resumes, the callback rates for applicants with a White-sounding name (column 1) an an African-American-sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these callback rates. In brackets in each cell is the number of resumes sent in that cell. Column 4 also reports the *p*-value for a test of proportion testing the null hypothesis that the callback rates are equal across racial groups.

Results

- Clear racial gap pattern
- For otherwise identical applications, white sounding names have a higher chance of receiving a call-back
- In overall sample
- For various subgroups
- Can we also classify employers?

Results

Table 2—Distribution of Callbacks By Employment Ad

Equal Treatment:	No Callback	1W + 1B	2W + 2B
88.13 percent	83.37	3.48	1.28
[1,166]	[1,103]	[46]	[17]
Whites Favored (WF):	1W + 0B	2W + 0B	2W + 1B
8.39 percent	5.59	1.44	1.36
[111]	[74]	[19]	[18]
African-Americans Favored (BF):	1B + 0W	2B + 0W	2B + 1W
3.48 percent	2.49	0.45	0.53
[46]	[33]	[6]	[7]
Ho: $WF = BF$			
$p = 0.0000$			

Results

- Large group of employers that treats equally (mostly driven by no call-backs at all)
- Significantly larger share of employers that prefer to call back whites compared to blacks...

Results

Driving forces

- Question: how do employers respond to the quality of the resume? Does response differ between black and white sounding names?
- Exploit variation in resume quality...

Results

Table 4—Average Callback Rates By Racial Soundingness of Names and Resume Quality

	Low	High	Ratio	Difference (p-value)
Panel A: Subjective Measure of Quality (Percent Callback)				
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)
Panel B: Predicted Measure of Quality (Percent Callback)				
	Low	High	Ratio	Difference (p- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

OLS Assumptions

OLS assumptions

OLS provides an appropriate estimator for the unknown regression parameters *if a set of assumptions* is true with respect to the linear regression model and the sampling scheme.

- Linearity in parameters.
- Sufficient variation in x_i.
- (x_i, y_i) are i.i.d.
- Zero conditional mean assumption.
- Homoskedasticity.

1 - Linearity

The linear regression model assumes that the relation between x and y (the population regression model) is **linear in its parameters** and correctly specified as $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

- OLS: β_0 is the intercept, β_1 the regression slope parameter.
- But what if the true relation is non-linear?
- How do we know if the assumption is plausible? Theoretical reasoning, data inspection, ...
- We will come back to this a bit. But non-linear regressions is mostly for more advanced lectures.

Question: Can you think of examples in which the relationship may not be linear?

2 - Sufficient variation

Recall the **univariate linear regression model**:
$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

- OLS minimises the sum of the squared residuals,

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

- The OLS estimators for β_0 and β_1 are:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{and} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

2 - Sufficient variation

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \ , \ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- If there is no (or little) variation in x_i, then $\sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$ (or is almost zero). In that case, $\hat{\beta}_1$ is not identified.
- We need the observed data x_i for $i = 1, ..., n$ to not be all of the same value, i.e. $var(X) > 0$.
- How do we know if the assumption is plausible? Data inspection!
- Something to consider to data collection! You will need variation in x

(Simulated) example

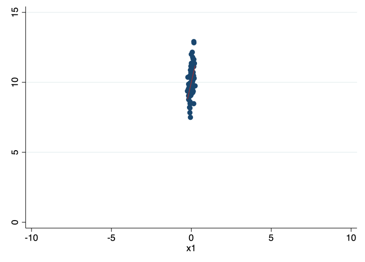
Generate random numbers: $x1$, $x2$, $x3$, y. xs have same correlation with y, same mean, but different standard deviation.

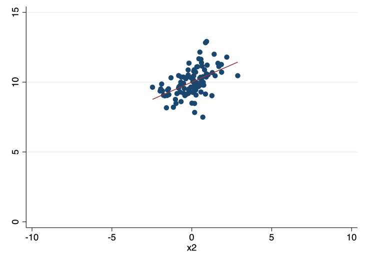
. corr
(obs=100)

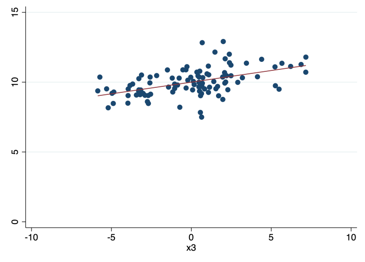
	x1	x2	x3	y
x1	1.0000			
x2	0.6000	1.0000		
x3	0.6000	0.6000	1.0000	
y	0.5000	0.5000	0.5000	1.0000

. summarize

Variable	Obs	Mean	Std. dev.	Min	Max
x1	100	5.48e-10	.1	-.2424136	.2378205
x2	100	4.48e-09	1	-2.459846	2.874251
x3	100	-1.99e-08	3	-5.853779	7.168859
y	100	10	10	-15.04229	39.11343







3 - (x_i, y_i) i.i.d.

OLS is based on the assumption that $(x_i, y_i),\ i = 1, ...n$ are **independently and identically distributed (i.i.d.)** across observations.

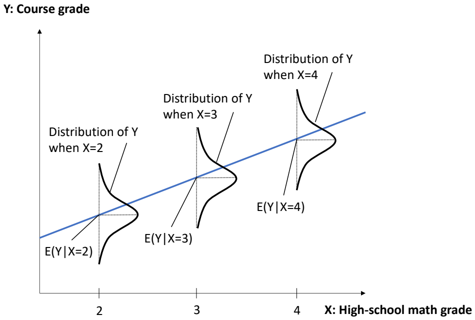
- **Random sampling:** (x_i, y_i) represent an i.i.d. random sample of size n following the population model.
- **Example:** A cross-sectional survey where individuals are randomly sampled from the population.
- **Counter-example 1:** Observations refer to the same unit of observation over time (time series data); here, observations correlate over time and are not independent.
- **Counter-example 2:** Independence can also be violated if observations are clustered (e.g. school classes).
- Important for data collection! Draw random samples!

4 - Zero conditional mean

The **zero conditional mean assumption** is:

$$\mathrm{E}[\epsilon_i | x_i] = 0$$

- "Other factors" contained in ϵ_i are unrelated to x_i: Given a value of x_i, the mean of the distribution of these is 0.



4 - Zero conditional mean

The **zero conditional mean assumption** is:

$$\mathrm{E}[\epsilon_i | x_i] = 0$$

- "Other factors" contained in ϵ_i are unrelated to x_i: Given a value of x_i, the mean of the distribution of these is 0.
- This condition implies mean independence, so no correlation between regressor and error term, i.e. $\mathrm{cov}(\epsilon, x) = 0$.
- **Randomised experiment:** Random assignment into treatment ($X = 1$) and control ($X = 0$) group makes x and ϵ independent - implies that conditional mean of ϵ given x is 0.
- **Without a randomised experiment,** we have to assume and argue that x is **as if randomly assigned**.
- This is a crucial assumption that we will return to repeatedly.

4 - Zero conditional mean

$$\mathrm{E}[\epsilon_i | x_i] = 0$$

- To see why this assumption is essential, recall that the estimator for β_1 can be written as:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \epsilon_i \left(x_i - \bar{x}\right)}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

- For the OLS estimator to be **unbiased** ($\mathrm{E}(\hat{\beta}_1) = \beta_1$), we need:

$$\mathrm{E}\left[\sum_{i=1}^{n} \left(x_i - \bar{x}\right) \epsilon_i\right] = 0$$

- $\mathrm{E}[\epsilon_i | x_i] = 0$ implies that $\mathrm{cov}(x_i, \epsilon_i) = 0$.
- Thus: If $\mathrm{E}[\epsilon_i | x_i] = 0$ holds, then $\mathrm{E}\left[\sum_{i=1}^{n} \left(x_i - \bar{x}\right) \epsilon_i\right] = 0$ and the OLS estimator is unbiased.

4 - Zero conditional mean

$$\mathrm{E}[\epsilon_i | x_i] = 0$$

- If the zero conditional mean assumption holds, we can further show that

$$\lim_{n \to \infty} \sum_{i=1}^{n} (x_i - \bar{x}) \, \epsilon_i = 0$$

- Thus, if the **zero conditional mean assumption** holds, the OLS estimator is both **unbiased**.
- **Important:** Do not confuse true error terms ϵ_i with regression residuals $\hat{\epsilon}_i$: $\hat{\epsilon}$ and x are uncorrelated by construction.

4 - Zero conditional mean

Key assumption for using OLS:

- If it holds: $\hat{\beta}_1$ is the (causal) **effect** of x on y.
- If not: $\hat{\beta}_1$ describes the **association** or **correlation** btw. x and y.

Note: Crucial assumption for interpretation (not for feasibility).

As long as there is enough variation in x, it will be possible to estimate an OLS regression. The important (and more challenging) part is to correctly interpret the results!

Strong assumption that often may not hold. How plausible is it?

4 - Zero conditional mean

- Again: $\mathrm{E}[\epsilon_i|x_i] = 0$ implies that $\mathrm{cov}(x_i, \epsilon_i) = 0$.

  $\mathrm{cov}(x_i, \epsilon_i) = 0$ does *not* necessarily imply $\mathrm{E}[\epsilon_i|x_i] = 0$,
  but $\mathrm{cov}(x_i, \epsilon_i) \neq 0$ implies that $\mathrm{E}[\epsilon_i|x_i] \neq 0$.

- We typically discuss the zero conditional mean assumption in terms of possible correlation between x_i and ϵ_i. Can we find reasons for which they should be correlated?

- The assumption rules out omitted variable bias that is that there are:

  unobserved confounders/third factors (unobserved variables correlated with x that also affect y), and/or

  reversed causality (x affects y and y affects x).

- *If unobserved confounders or reversed causality are likely: the assumption is probably violated!*

5 - Homoskedasticity

With the assumption so far, we know that the sampling distribution of $\hat{\beta}_1$ is centred around the population parameter β_1, but we do not know the population variance.

- To derive the variance of OLS estimators: assume **homoskedasticity** + **no correlation**

$$\mathrm{var}(\epsilon_i|x_i) = \sigma^2 \qquad \text{and} \qquad \mathrm{cov}(\epsilon_i, \epsilon_j) = 0 \quad i \neq j$$

- **Homoskedasticity**: constant error-term-variance across observations.
- If assumption of homoskedasticity is violated: **heteroskedasticity**.

  → Does *not* bias OLS estimates, but leads to incorrect standard errors and potentially misleading inference.

5 - Heteroskedasticity

If assumption of homoskedasticity is violated: **heteroskedasticity**.

- **Heteroskedasticity**:

$$\mathrm{var}(\epsilon_i | x_i) = \sigma_i^2$$

- Interpretation: variance of error terms differs between observations.
- Example: Variation in annual earnings tends to be larger for higher than for lower educated (bounded by minimum wage).
- Use **robust** standard errors (White-Huber) to adjust for heteroskedasticity.

5 - No correlation

If assumption of no correlation in error terms is violated:

$$\text{cov}(\epsilon_i, \epsilon_j) \neq 0 \text{ for some } i \neq j$$

- Often the case in **time-series data** (i.e., i and j are two time periods), or
- when observations are **clustered** (e.g., students of the same class, individuals of the same neighbourhood, workers of the same firm).
- Possible to adjust standard errors.
- Most statistical softwares easily allow to use robust standard errors and/or clustered standard errors.

OLS assumptions

Summary of key assumptions for OLS:

- **Linearity in parameters**: Otherwise our model is wrong.
- **Sufficient variation in x_i**: Otherwise estimation fails.
- (x_i, y_i) **are i.i.d.**: Ensures independence and identical distribution of observations.
- **Zero conditional mean assumption**: Necessary for causal interpretation of slope parameters.
- **Homoskedasticity**: Needed for standard errors (else use heteroskedasticity robust standard errors).

Multivariate Regression

Control Variables in Regressions

- What can we do if zero conditional mean assumption does not hold?
- Third factor / omitted variable bias may be (at least partially) addressed by including the third factors/ omitted variables in the regression as control variables
  - Works of course only if these variables are observable!
  - Often, third factors / omitted variables are unobservable (e.g., IQ, motivation. . .)

Multivariate Regressions

- Suppose there is an outcome Y with observed realization $(y_1, \cdots, y_N)$
- And two independent variables (predictors) X_1 and X_2 with observed realization $(x_{11}, \cdots, x_{1N})$ and $(x_{21}, \cdots, x_{2N})$
- Linear regressions assume that

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

- i is the unit of observation
- β_0 is the intercept
- β_1 and β_2 the marginal effects (slope) wrt X_1, X_2 respectively
- ϵ_i is the error term
- x_2 is sometimes referred to as a control variable.

Interpretation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

- As in the univariate case, the interpretation of the coefficients hinges on the **Zero Conditional Mean Assumption**:

$$\mathrm{E}[\epsilon_i | x_i] = 0 \qquad \text{becomes} \qquad \mathrm{E}[\epsilon_i | x_{1i}, x_{2i}] = 0$$

- If the assumption holds, OLS estimators are **unbiased**.
- In this case, the estimated effects can be interpreted as **causal** effects (**ceteris paribus**).

Interpretation

Recall interpretation of β_1 in the **univariate linear regression model**:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- The coefficient β_1 is the **partial** or **marginal effect** $\partial \mathrm{E}[Y|\mathbf{x}]/\partial \mathbf{x}$.
- Answer to: How much does y change if x changes by one unit?

In the model with two regressors:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

- Interpretation of β_1: **the marginal effect** $\partial \mathrm{E}[y_i|x_{2i}]/\partial x_{1i}$.
- **Ceteris paribus interpretation:** The marginal effect of x_1 keeping x_2 the same (intuitively: in subsample with same value of x_2).
- Answer to: How much does y change if x_1 changes by one unit conditional on the observed value of x_2?

Control Variables in Regressions

- If we are interested in the effect of x_1 (education) on y (wages) holding x_2 (IQ) constant, we can use multivariate regressions.
- We need data on all of these objects
- A multivariate OLS delivers the best approximation in the sense of minimizing the squared deviations.

Variable selection

Data sets often contain many variables and we have to choose.

- Omitting relevant variables may cause violation of zero conditional mean assumption (more later).
- Including too many variables, however, may also be problematic:
  1. **Post-treatment variables, mediators or mechanisms**
     * "Controlling" for variables that can be thought of as the causal paythway between x_1 and y may block the effect of x_1 on y that you are trying to measure
     * Hence: do not include variables through which the treatment effect of x_1 on y may likely operate
     * causal path: $x_1 \rightarrow x_2 \rightarrow y$

Variable selection

Data sets often contain many variables and we have to choose.

- Omitting relevant variables may cause violation of zero conditional mean assumption (more later).
- Including too many variables, however, may also be problematic:

  2. **Bad controls or coliders**
     * A collider is a variable that is caused by explanatory variable x_1 and outcome variable y. Assume x_2 is the collider. Controlling for x_2 may introduce bias in the estimation of the potential causal effect of x_1 on y because it leads to a non-causal relationship (correlation) between x_1 and y.
     * Hence: do not include variables that may be colliders
     * causal paths: $x_1 \rightarrow x_2 \leftarrow y$

Variable selection

So, how do we choose which variables to include in the regression?

- **Use theory, own thinking and common sense to decide which variables should be included! $\rightarrow$ Task of the data analyst.**
- Not in this course: Use machine learning techniques to help with variable selection.

Frankfurt School of Finance & Management
German Excellence. Global Relevance.

Data Collection

Experimental design

Frederik Schwerter

Frankfurt School

Autumn, 2023

So far

- Naturally occurring data may not necessarily allow to study causal relationships
- Crucial: zero condition mean assumption! OLS uncovers causal effect of x on y only if x is independent of the error term
- Zero condition mean assumption is violated due to omitted variables (often also called third factors) that are part of the error and are correlated with x
- If the omitted variables that violate the zero conditional mean assumption are observable, including them in multivariate regression analyses allows to control for them: the causal effect of x may then be estimatable
- But: it is often unclear whether all omitted variables are observable
- And: including control variables may also introduce bias in the first place (colliders, mediators)
- More general solution: random assignment of treatments, like in an experiment

Today

- The 101 of experimental design
- Experiments are often called RCT: randomized controlled trails
- The explanatory variable is oftentimes called “treatment”

What is an experiment?

- Encyclopedia: “A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a fact"
- Different than with naturally occurring data, experimentalists are directly involved in the data-generating process
- Types of experiments:
  - Lab: artificial but lots of control
  - Lab in the field: participants (partly) know being in an experiment
  - Field: experimentalist controls/affects treatment assignment
  - “Natural" Field: exploit naturally occurring (random) treatments.

Which experiment should I use?

- The different types of experiments have different advantages and disadvantages

- Which experiment would you want to run in your business?
  Depends on the question that you want it to speak to!

- Try to condense your question into somethings simple/straightforward that is potentially answerable:

  Yes: Can changing management practice ABC increase measurable productivity of employees subject to the practice?

  No: Can changing management practice ABC turn company into the biggest company in the world at some point in the future?

What experiments can do? Gain control over:

- Treatment

- Assignment

- Measurement

Treatment 1/

- Treatment: a particular condition of the experiment
- Control: often separately named, but just another condition of the experiment, typically where "the treatment" is removed/inactive/reduced
- Often a RCT compares the treatment with a control; but there could be more treatments
  - In some RCTs, one may only have several treatments (no control) and compare treatments to each other

Treatment 2/

- Between treatment and control(/ or other treatments) one tries to change only one thing (at best) or as few things as possible, everything else is constant.
  - Note: the differences in the design of treatment and control could serve as explanations for potential differences in the outcomes of treatment and control. This implies that it is imprecise what the actual cause of a potential causal effect is. To gain maximal control over this interpretation: limit the difference in design as much as possible
  - Think: medical trials, where the control group receives placebos to keep the experience of participants as equal as possible to participants of the treatment group

Treatment 3/

- What treatment manipulation should you use?
- This comes directly from the research question
  - It you want to know whether a training program affects performance, then you should simply vary whether people participate in the program or not: the treatment is the program
- Sometimes it is a bit more tricky, but it typically depends on the precise research question. But you can even have vary subtle treatment manipulations:
  - Priming: make a particular experiences/identity dimensions salient to activate particular character properties

Treatment 4/

- Mechanism: why does a particular treatment work?
- Often, we may not necessarily know this directly from a treatment intervention because the treatment intervention is a collection of things that are different to a control condition
  - For instance, a training program may teach special techniques of how to improve performance
  - Or you may simply form greater bonds with your coauthors and enjoy working more because you feel you work together as a team
  - Knowledge of the mechanism would allow for a more targeted/cost-efficient program. Eg, maybe only regular social events are required, and not “expensive” teachers for training programs

Assignment

- Treatments should be assigned randomly!
- By definition this achieves that the zero conditional mean assumption holds
- One can then use OLS to study the causal effect of the treatment on the outcome variable

Treatment assignment: stratification

- Goal of randomization: treatment groups should be on average the same
- To enforce this: employ a stratified randomization
- E.g., to ensure balance between treatment and control in work experience, form pairs of participants with the same work experience and randomize treatment assignment within pairs
- Stratification on multiple variables (e.g., wage, gender, age, ...) is also possible, but technically more demanding
- Important issue: stratification is only possible for observable variables that you need access to during randomization!

- Statistical consequences of stratification: stratification changes how standard errors of the estimated treatment effect should be calculated
- A detailed account of this is beyond the scope of this class
- Note though that stratification allows for weakly smaller standard errors and hence implies more precise inference
- Advanced literature: Athey, Susan, and Guido W. Imbens. “The econometrics of randomized experiments." Handbook of economic field experiments. Vol. 1. North-Holland, 2017. 73-140.
- R package for stratified randomization: “randomizr"

- In designing your experiment, always anticipate the type of statistical analysis you will want to do to evaluate the exp
  - Typically comparison of means between treatment and control. But how many independent observations do you have? This will affect standard errors and hence how precise/powerful your tests are
- Difference between observations and statistically independent observations: e.g., if five employees work in a team, their individual output cannot count as independent observations, instead only the team's output is an independent observation

Measurement

- You may want to add measurements, like surveys or tests, beyond naturally occurring outcomes, like sales at the end of the year, to have more direct evidence that allows to evaluate your experiment

- Recall CFM: their outcome was “merely a lab measure on lying"

- Actually: experiments can be only about measurement!

  Also note, already some companies use employee surveys to learn beyond traditional outcomes

- Advantage: you can observe things that are unobservable in naturally occurring data

  - Beliefs, psychological traits, implicit and explicit preferences, physiological responses, attention

- For instance, assume you are worried about discrimination at work, you could use the implicit association test

  Lets do one now, click on your respective language preference: German or English

- Qualitative vs quantitative measures?
  - Tradeoff:
  - qualitative is often easier to understand, but interpersonal comparability is less clear
  - quantitative requires more time and focus of participants, but is clearer to interpret
- Hypothetical vs incentivized measures?
  - Tradeoff:
  - hypothetical is often easier to understand, but may be biased towards socially desirable responses
  - incentivized: requires participants understanding/trust, but more likely elicits individuals' true belief, preference, attitude; important though: individuals are actually incentivized to report truthfully (e.g.: difference between first and second price auction)

Concerns/Considerations

- Third factor / omitted variable threat
  - Feedback/spillover effects
  - Hawthorne effect
  - Experimenter demand
- External validity
- Convenience sample
- Statistical power

Feedback/spillover effects

- What if the treatment spills over to the control? Eg, the employees who did the training program taught employees who didnt participate the tricks of the training program. Comparing treatment and control may underreport the true causal effect.
- Be aware of the incentive structure at work, what information employees have, how interactions are structured and what are contextual features
- Remedy: eg, randomize at location level of a company, not within locations

Hawthorne effect

- Individuals modify an aspect of their behavior in response to their awareness of being observed
- History: at Hawthorne Works, a Western Electric factory, worker productivity increased due to minute increases in illumination, but only for a short time. Workers knew they are part of study
- Recent debate: original results may not provide evidence for Hawthorne effect — yet, researchers are still concerned about the effect
- Potential solution: design control to include inconsequential changes such that potential Hawthorne effects are constant between treatment and control (note similarity of placebos in medical trials)

Experimenter demand

- Experimental artifact where participants form an interpretation of the experiment's purpose and subconsciously change their behavior to fit that interpretation

- Solution: in any instructions, keep language neutral and give participants are little information as possible about your intentions

- Like for Hawthorne effect: try to include inconsequential changes also in the control such that potential experimenter demand effects are constant between treatment and control

External validity

- Results of experiment may be internally valid, e.g., causal effects are reasonably identified
- Results, however, may not be externally valid, that is, in other contexts they would not show up
- E.g., experiment was conducted in only a few locations and these locations may be different (in terms of employees, economic conditions, incentives, ...) from the other locations of the company. Results of the experiment may hence not apply to the other locations
- The more lab-like the experiment, the more is external validity a concern; but even natural field exp. face the concern

- What can be done about external validity?
- Try to design experiment such that participating units are representative
- E.g., when you test whether an anti-discrimination training increases performance in a continent-wide company, the experiment should cover (both treatment and control) employees from all relevant countries in proportional size

Convenience sample

- Sometimes, not everyone participates in experiments/surveys when they are voluntary
- If we do not have a good understanding of who selects into experiment (that is, into treatment and control), we may have only access to a convenience sample
- Even if treatment and control group are on average very similar due to randomization, the experimental sample (treatment + control) may not be representative for their population: external validity concerns!
- Potential solution: low entry costs for potential participants

Statistical power

- Has the intended experiment a reasonable chance of finding results of the size that one might reasonably expect?
- Such questions depend on a number of inputs, and can focus on various outputs
- Assume you know how large your treatment and control groups will be ($+$ how “powered" your test shall be and what the significance level will be, conventionally 80% and 5% respectively)
- Then you could calculate minimal detectable effect size d,

  where $d = \frac{\Delta}{SE}$ with SE being the standard error

  is the standardized treatment effect

- in R you can use, e.g., package pwr
- Function pwr.t2n.test(n1 $=$ NULL, n2 $=$ NULL, d $=$ NULL, sig.level $=$ NULL, power $=$ NULL, alternative=c("two.sided")) can be used for two sample t-tests
- E.g.: assume you have a treatment group of 50 and a control group of 100 employees and you apply conventional levels for power and significance level,
- then pwr.t2n.test(n1 $=$ 50, n2 $=$ 100, d $=$ NULL, sig.level $=$ 0.05, power $=$ 0.8, alternative=c("two.sided")) calculates a minimal detectable effect size of $d = 0.488$

- Your sample size grants you a reasonable chance of detecting an effect size of $d = 0.488$

- According to Cohen (1988, Statistical power analysis for the behavioral sciences):
  - $d = .1$ refers to small effects
  - $d = .25$ refers to medium effects
  - $d = .4$ refers to large effects

- So: in case of above example, the sample size allows “only" reasonable chance to find large effects; which implies that the experiment has no/little reasonable chance to detect small or medium effects

- But maybe even small/medium effects could be economically relevant for a company: problem!

- Designing well-powered experiments should hence be the goal!

  How?

- Before designing the experiment, try to figure out the smallest effect size such that the effect would be meaningful

- Determine the minimum sample size to run a well-powered experiment given the smallest effect size that is meaningful

- Design the experiment to include as many independent observations as you determined in the earlier step

- If you cannot “deliver" the required sample, running the experiment ***may*** not be a useful exercise!

Case study

- Cohn, Fehr, Marechal: “Business culture and dishonesty in the banking industry” (2014, Nature)