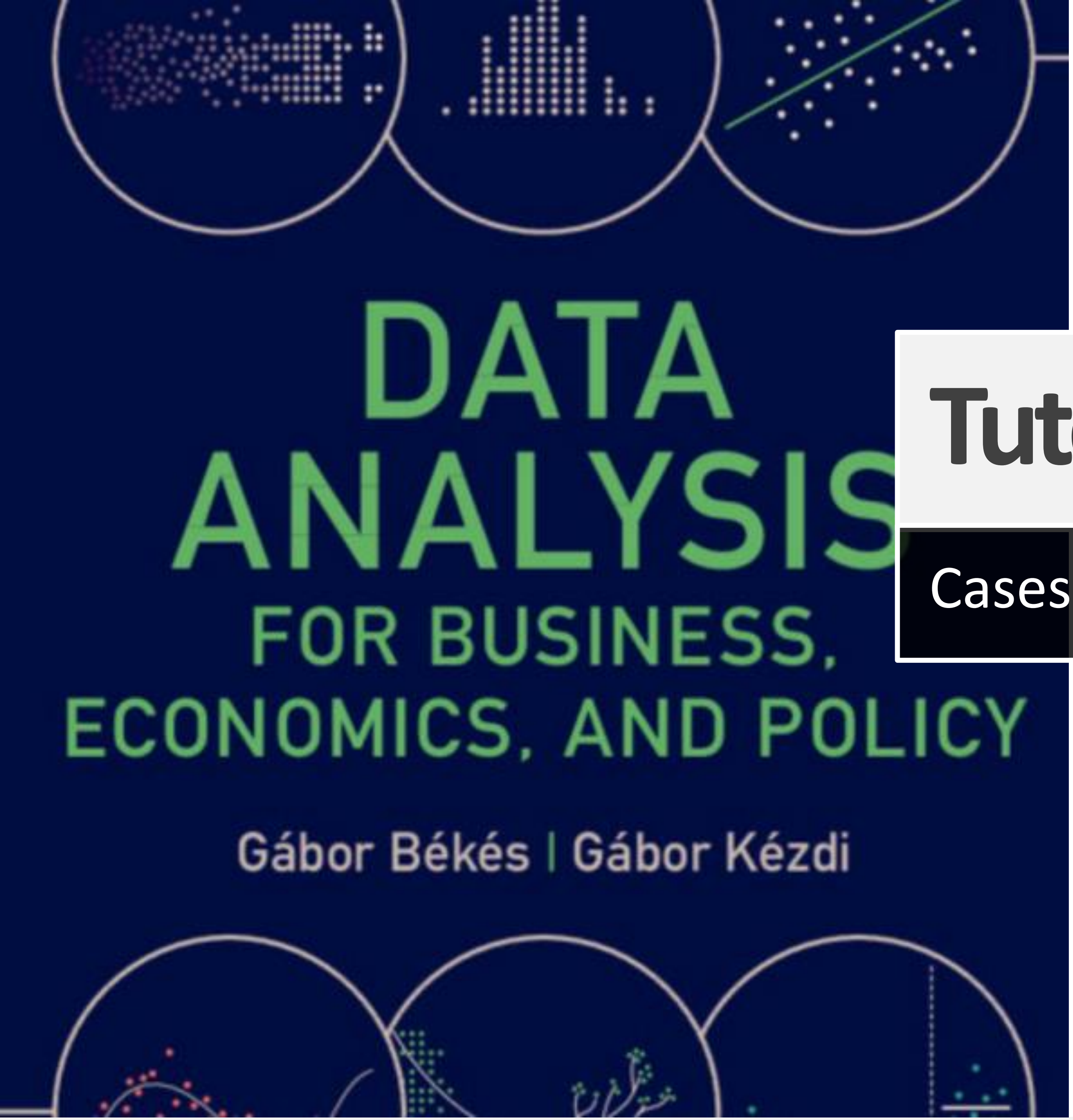

Lecturer: Attila Vig

Tutorials

Cases from the book accessible online

In the tutorials we use cases from the book

& we also use real financial data to gain in-depth understanding for applied finance research work, relevant for the industry

Issues/Problems

- If you had any issues with any of the previous data, please ask during the tutorial or post on Moodle in the relevant week discussion forum

Tutorial (Week 5)

- Wednesday:
  Airline merger data, data cleaning, understanding data

  Data

  Code, R

  Creating variables, intro to Diff-in-Diffs

- Friday
  - Diff-in-Diff analysis more in depth

  Code, R

The Case is Airline mergers

- Many things happen at airline mergers
- Company may become more profitable
- Company may enjoy more monopolistic position
  - Company may be able to charge higher prices.

This is what we examine now ☺

- Company become less profitable as the company becomes too big, airline industry is unionized in the US and quite difficult to fire staff, often resulting in inefficiencies and strikes.

The Case is Airline mergers - causality

- Y (price change) = a+ b*Airline merger +c* Controls + e
- Y – price change happens over time, simply the effect of inflation
- Merger is a time dummy, so yes over time price increases most likely but not because of the merger
- Let’s try to sort out the data.

How can we “nail down” causality?

We will use a difference in differences approach, but the first thing is actually get a data, where we can focus on measuring effect.

The Case is Airline mergers - causality

- Airline ticket pricing is for specific routes, different, so we first have to disentangle market, identify tickets bought for the same routes before and after the merger
- We also have to differentiate across tickets for flights on companies involved in the merger, versus the competition flights

The Case is Airline mergers – data preparation

- It is very important to be able to handle effectively data
- If you cannot prepare the data in way to be able to work with it, run regressions, analysis, test mean difference, then everything is lost.
- So, please always spend some time on thinking about data, data structure and how you need to arrange the data before blindly trying to run regressions.
- This is a messy data. (a realistic real world example)

The Case is Airline mergers – data preparation

- Creating routes, before that make sure correctly pool in the data and prepare

```
# load theme and functions
source("ch00-tech-prep/theme_bg.R")
source("ch00-tech-prep/da_helper_functions.R")
options(digits = 3)

data_in <- paste(data_dir,"airline-tickets-usa","clean/", sep =
"/")
use_case_dir <- "ch22-airline-merger-prices/"


data_out <- use_case_dir
output <- paste0(use_case_dir,"output/")
create_output_if_doesnt_exist(output)
```

The Case is Airline mergers – data preparation

```
# CREATE Workfile : only before and after period
data <- read_dta(file.path(data_in, "originfinal-panel.dta"))
# from OSF
#data <- read_dta("https://osf.io/zw2h9/download")
# ************************************************************************
# * market = origin X final destination
# * (note final destination is:
# * airport at end of one-way routes if 4 or fewer
# * airport in middle of return routes if there is middle & 9 or fewer
# * before = 2011 (all year)
# * after  = 2016 (all year)
# * workfile 1: drop all other years
data <- data %>%
  filter(year==2011 | year==2016)
```

The Case is Airline mergers – data preparation

```
# * create total number of passengers from shares
# * so we can get aggreagate shares
data_agg <- data %>%
  mutate(
    ptotalAA = shareAA*passengers,
    ptotalUS = shareUS*passengers,
    ptotallargest = sharelargest*passengers
  ) %>%
  group_by(origin, finaldest, return, year) %>%
  summarise(
```

Code continues on next slide

The Case is Airline mergers – data preparation

```
airports = first(airports),
    return_sym = first(return_sym),
    stops = first(stops),
    ptotalAA = sum(ptotalAA),
    ptotalUS = sum(ptotalUS),
    ptotallargest = sum(ptotallargest),
    passengers = sum(passengers),
    itinfare = sum(itinfare)
  ) %>%
  ungroup()
```

The Case is Airline mergers – data preparation

```
data_agg <- data_agg %>%
  mutate(
    after = as.numeric(year == 2016),
    before = as.numeric(year == 2011),
    avgprice = itinfare/passengers,
    shareAA = ptotalAA/passengers,
    shareUS = ptotalUS/passengers,
    sharelargest = ptotallargest/passengers,
    AA = as.numeric(shareAA > 0), #share variables never missing
    US = as.numeric(shareUS > 0),
    AA_and_US = as.numeric(shareAA > 0 & shareUS > 0),
    AA_or_US = as.numeric(shareAA > 0 | shareUS > 0)
  )
```

The Case is Airline mergers – data preparation

```
#create numeric ID for market
data_agg <- data_agg %>%
  arrange(origin, finaldest) %>%
  mutate(market = factor(paste(origin, finaldest, return, sep = "_")))
# passengers before and after
data_agg <- data_agg %>%
  arrange(market, year) %>%
  group_by(market) %>%
  mutate(
    pass_bef = mean(ifelse(before == 1, passengers, NA), na.rm = TRUE),
    pass_aft = mean(ifelse(after == 1, passengers, NA), na.rm = TRUE)
  ) %>%
  ungroup()
```

The Case is Airline mergers – data preparation

```
# * balanced vs unbalanced part of panel
data_agg <- data_agg %>%
  group_by(market) %>%
  mutate(balanced = as.numeric(n() == 2)) %>%
  ungroup()


data_agg %>%
  group_by(balanced) %>%
  summarise(sum(passengers), n())
```

The Case is Airline mergers – data preparation

```
# Define treated and untreated markets
# treated: both AA and US present in the before period
# untreated: neither AA nor US present in the before period OR only AA or only US in before period
# NOTE There is a error in the book on p628:
# Original version with error:
# This definition of treated and untreated markets left some markets neither treated
#  nor untreated: those with only American or only US Airways present in 2011. For the
#  main analysis we **dropped these from the data**.

# Corrected:
# This definition of treated and untreated markets left some markets neither treated
#  nor untreated: those with only American or only US Airways present in 2011. For the
#  main analysis we **kept these in the data as untreated**.
```

The Case is Airline mergers – data preparation

```
data_agg <- data_agg %>%
  arrange(market, year) %>%
  group_by(market) %>%
  mutate(
    treated = ifelse(balanced == 1, max(as.numeric(AA_and_US == 1 & before == 1)), NA),
    untreated = ifelse(balanced == 1, max(as.numeric(AA_or_US == 0 & before == 1)), NA),
    smallmkt = max(as.numeric(passengers < 5000 & before == 1))
  ) %>%
  ungroup()
data_agg <- data_agg %>%
  mutate(lnavgp = ifelse(is.infinite(log(avgprice)), NA, log(avgprice))) %>%
  arrange(market, year) %>%
  group_by(market) %>%
  mutate(d_lnavgp = lnavgp - lag(lnavgp)) %>%
  ungroup()
```

The Case is Airline mergers – data preparation

```
 ***************************************************************************
# * DESCRIBE
# describe yearly data
data_agg %>%
  select(year, passengers) %>%
  group_by(year) %>%
  summarise_each(funs(N = length,
                      q25 = quantile(., 0.25),
                      median = median,
                      q75 = quantile(., 0.75),
                      q90 = quantile(., 0.90),
                      mean = mean,
                      sum = sum)
                    )
```

The Case is Airline mergers – data preparation

```
data_agg %>%
  filter(origin=="JFK" & finaldest=="LAX") %>%
  select(market, origin, finaldest, return, year, passengers)
data_agg %>%
  filter(year == 2011) %>%
  select(smallmkt, passengers) %>%
  group_by(smallmkt) %>%
  summarise_each(funs(N = length,
                      min = min,
                      max = max,
                      median = median,
                      mean = mean,
                      sum = sum))
```

The Case is Airline mergers – data preparation

```
# describe balanced
data_agg %>%
  group_by(year, balanced) %>%
  summarise(n(), sum(passengers), mean(passengers))


# describe treatment
data_agg %>%
  group_by(year, treated, untreated) %>%
  summarise(n(), sum(passengers), mean(passengers))
```

The Case is Airline mergers – data preparation

```
# describe outcome
data_agg %>%
  filter(before == 1) %>%
  select(avgprice) %>%
  summarise_all(funs(N = length, ……
…….))
data_agg %>%
  filter(avgprice==0) %>%
  select(passengers) %>%
  summarise_all(funs(N = length,
                     mean = mean,
                     sum = sum))
write_rds(data_agg,paste0(data_out,"ch22-airline-workfile.rds"))
```

Summary stat tables should include at minimum what is here in the code and in the book:

```
min = min,
            max = max,
            median = median,
            mean = mean,
            sum = sum
```

But, I and the world is now working often with quantile regression, good to report quartile or quantile values

End

NOTE:

Now, you can see messiness of data, most of the time, data cleaning 70-80% of the time.

To get the data in a shape you can and able to work with, takes long time. Do not underestimate that in your work, in your studies.

The Case is Airline mergers – Analysis

- Let’s deep dive into the Airline example again

- Wednesday, we tried to understand cleaning of the data,
- Preparing to be able to do a Difference and difference analysis

Code, R

The Case is Airline mergers – Analysis

- Let’s go through the code again, and replicate the book analysis

- Make sure you are also able to make figures, and understand the work