GV300 - Quantitative Political Analysis

University of Essex - Department of Government

Lorenzo Crippa

Week 24 – 9 March, 2020

2020 Department of Government Student Conference

Teaching evaluation. Be mindful of your unconscious biases!

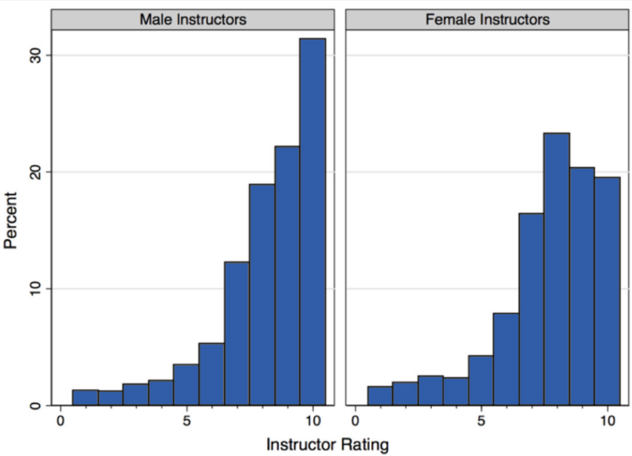


Figure 1: Source: Rivera and Tilcsik 2019

Today's class

1. Prep for exam: From Mac (or Linux) to PC
2. Correction of Problem Set 7
3. Maximum Likelihood Estimation

Prep for exam: From Mac (or Linux) to PC

RStudio and Stata between Mac and PC

- RStudio and Stata work identically on Mac (or Linux) and PC.

RStudio and Stata between Mac and PC

- RStudio and Stata work identically on Mac (or Linux) and PC.
- You only need to change directory paths because Mac and Linux are UNIX systems, while Microsoft Windows is NT-based.

RStudio and Stata between Mac and PC

- RStudio and Stata work identically on Mac (or Linux) and PC.
- You only need to change directory paths because Mac and Linux are UNIX systems, while Microsoft Windows is NT-based.
- These systems treat the / and \characters differently, and we use these characters in setting up paths (for working directories, for saving graphs, for getting data...)

RStudio and Stata between Mac and PC

- RStudio and Stata work identically on Mac (or Linux) and PC.
- You only need to change directory paths because Mac and Linux are UNIX systems, while Microsoft Windows is NT-based.
- These systems treat the / and \characters differently, and we use these characters in setting up paths (for working directories, for saving graphs, for getting data...)
- In a UNIX system (Mac or Linux) the / is different from a \. you have: `/Users/Your_name/Documents/your_folder`

RStudio and Stata between Mac and PC

- RStudio and Stata work identically on Mac (or Linux) and PC.
- You only need to change directory paths because Mac and Linux are UNIX systems, while Microsoft Windows is NT-based.
- These systems treat the / and \characters differently, and we use these characters in setting up paths (for working directories, for saving graphs, for getting data. . .)
- In a UNIX system (Mac or Linux) the / is different from a \. you have: `/Users/Your_name/Documents/your_folder`
- In Microsoft Windows the / is identical to a \. So you can have both: `C:/Users/Your_name/Documents/your_folder` and `C:\Users\Your_name\Documents\your_folder`

Dealing with two different systems

If you need to work on both systems in R or Stata you can include both paths and comment out the one you don't need (depending on the machine you're working on).

Dealing with two different systems

If you need to work on both systems in R or Stata you can include both paths and comment out the one you don't need (depending on the machine you're working on). My examples:

Dealing with two different systems

If you need to work on both systems in R or Stata you can include both paths and comment out the one you don't need (depending on the machine you're working on). My examples:

```
# Lorenzo Macbook
setwd("/Users/Lorenzo/Dropbox/Shared_Essex/GTA/GV300")

# Lorenzo Essex
#setwd("C:/Users/lc19059/Dropbox/Shared_Essex/GTA/
    GV300")
```

```
* Lorenzo Macbook
* cd "/Users/Lorenzo/Dropbox/Shared_Essex/GTA/GV300"

* Lorenzo Essex
cd "C:/Users/lc19059/Dropbox/Shared_Essex/GTA/GV300"
```

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do
- Open the working directory in File Explorer

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do
- Open the working directory in File Explorer
- Copy the path of your directory (the text in the bar that you have on top of the File Explorer)

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do
- Open the working directory in File Explorer
- Copy the path of your directory (the text in the bar that you have on top of the File Explorer)
- Paste it in RStudio or Stata under the respective “change directory” command `setwd()` or `cd`

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do
- Open the working directory in File Explorer
- Copy the path of your directory (the text in the bar that you have on top of the File Explorer)
- Paste it in RStudio or Stata under the respective "change directory" command `setwd()` or `cd`
- R works only if you turn each \into a / while Stata works both ways

Practical tips for the exam

The exam is going to be on PCs. Follow these steps:

- Set the working directory as a FIRST thing you do
- Open the working directory in File Explorer
- Copy the path of your directory (the text in the bar that you have on top of the File Explorer)
- Paste it in RStudio or Stata under the respective "change directory" command `setwd()` or `cd`
- R works only if you turn each \into a / while Stata works both ways
- In any case it's always better to change each \into a /

Correction of Problem Set 7

Question 1 – (a)

Difference-in-differences: generate indicators for treatment-control groups and pre-post intervention. List countries in the two groups.

Question 1 – (a)

Difference-in-differences: generate indicators for treatment-control groups and pre-post intervention. List countries in the two groups.

```
data$treatment <- ifelse(data$yearJoinEU == 2004, 1, 0)
data$intervention <- ifelse(data$year >= 2004, 1, 0)
```

```
gen treatment = 0 + (yearjoineu == 2004)
gen intervention = 0 + (year >= 2004)
```

Question 1 – (a)

Difference-in-differences: generate indicators for treatment-control groups and pre-post intervention. List countries in the two groups.

```
data$treatment <- ifelse(data$yearJoinEU == 2004, 1,
    0)
data$intervention <- ifelse(data$year >= 2004, 1, 0)
```

```
gen treatment = 0 + (yearjoineu == 2004)
gen intervention = 0 + (year >= 2004)
```

Countries in the treatment group: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic, Slovenia.

Question 1 – (a)

Difference-in-differences: generate indicators for treatment-control groups and pre-post intervention. List countries in the two groups.

```r
data$treatment <- ifelse(data$yearJoinEU == 2004, 1,
    0)
data$intervention <- ifelse(data$year >= 2004, 1, 0)
```

```stata
gen treatment = 0 + (yearjoineu == 2004)
gen intervention = 0 + (year >= 2004)
```

Countries in the treatment group: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic, Slovenia.

Countries in the control group: Albania, Armenia, Bulgaria, Croatia, Georgia, Kosovo,Macedonia, FYR Moldova, Montenegro, Serbia, Ukraine

Question 1 – (b)

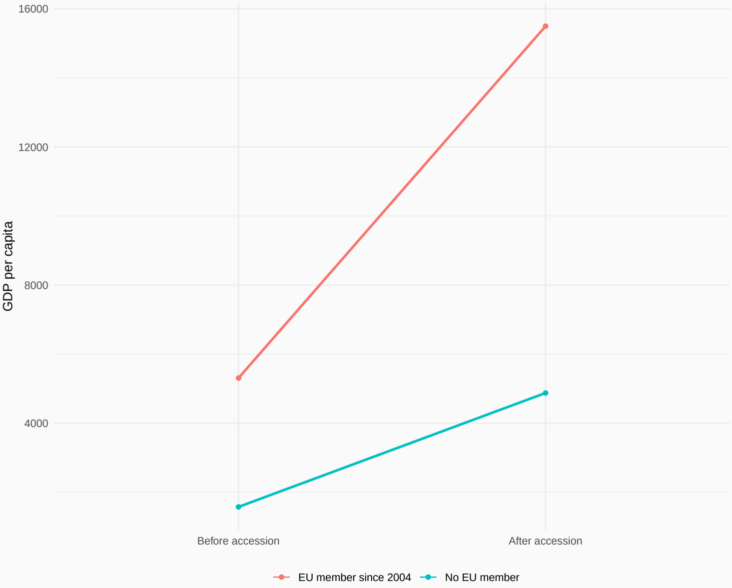
Compute the mean of `GDPPerCapita` for treatment and control group pre- and post-intervention. Plot those numbers. Compute the difference-in-differences from those numbers.

Question 1 – (b)

Compute the mean of `GDPPerCapita` for treatment and control group pre- and post-intervention. Plot those numbers. Compute the difference-in-differences from those numbers.

The difference in differences is
$(15495.61 - 5306.678) - (4873.54 - 1574.61) = 6890.301$

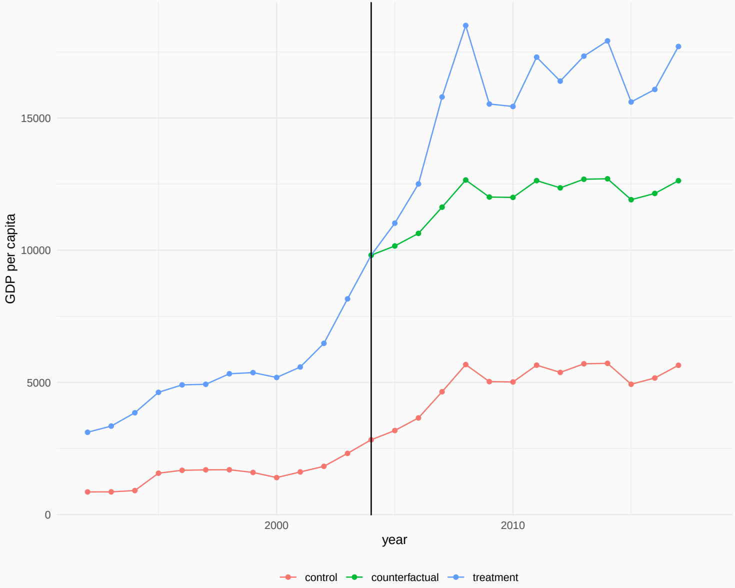
Question 1 – (b)


16000
12000
8000
4000
GDP per capita
Before accession
After accession
EU member since 2004
No EU member

Question 1 – (c)

Plot the mean of `GDPPerCapita` for treatment and control group over `year`. Add a line indicating the intervention year. Add the counterfactual `GDPPerCapita`. Evaluate whether the common trend assumption is met pre-intervention. Are the parts of SUTVA met that are relevant to the goodness of the differences-in-differences estimator? Why or why not?

Question 1 – (c)



15000
10000
5000
0
GDP per capita
2000
2010
year
control
counterfactual
treatment

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group.

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group
2. no spill-over effects

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group
2. no spill-over effects

- Treatment group is constant: no country that joined the EU in 2004 left the EU until the end of the data set.

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group
2. no spill-over effects

- Treatment group is constant: no country that joined the EU in 2004 left the EU until the end of the data set.
- Some countries in the control group joined the EU later which may bias our estimate of the causal effect of joining the EU.

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group
2. no spill-over effects

- Treatment group is constant: no country that joined the EU in 2004 left the EU until the end of the data set.
- Some countries in the control group joined the EU later which may bias our estimate of the causal effect of joining the EU.
- Spill-overs: it certainly may be that a rising GDP per capita thanks to joining the EU in, say, Slovenia may have affected the GDP per capita in neighbouring Croatia.

Question 1 – (c): SUTVA

It seems that GDP per capita is running in parallel in treatment and control group. SUTVA for DiD requires:

1. stability in composition of treatment and control group
2. no spill-over effects

- Treatment group is constant: no country that joined the EU in 2004 left the EU until the end of the data set.
- Some countries in the control group joined the EU later which may bias our estimate of the causal effect of joining the EU.
- Spill-overs: it certainly may be that a rising GDP per capita thanks to joining the EU in, say, Slovenia may have affected the GDP per capita in neighbouring Croatia.

Hence, SUTVA may be violated

Question 1 – (d), (e) and (f)

(d) Run a regression to compute the differences-in-differences estimator. Report and interpret the result. Speak to the three relevant coefficients.

(e) Improve your regression in (d) by computing clustered standard errors.

(f) Improve your estimate of the causal effect of joining the EU in (e) by including one relevant country-level covariate into the regression. Report and interpret your result. Was your estimate in (e) an over- or underestimate of the causal effect? Speculate why failing to include this covariate led to bias in your estimate in (e).

Question 1 – (d), (e) and (f)

	(d)	(e)	(f)	(f)
(Intercept)	1574.91^{***}	1574.91^{***}	-497.75	6153.33^{***}
	(299.71)	(136.46)	(314.89)	(588.39)
intervention	3298.63^{***}	3298.63^{***}	2950.32^{***}	3361.34^{***}
	(390.18)	(286.10)	(297.74)	(225.31)
treatment	3731.77^{***}	3731.77^{***}	2961.21^{***}	241.61
	(451.94)	(315.27)	(349.38)	(547.94)
intervention:treatment	6890.30^{***}	6890.30^{***}	5751.81^{***}	6827.59^{***}
	(593.70)	(575.16)	(608.09)	(547.98)
exportsShareGDP			64.37^{***}	
			(9.00)	
yearJoinEU				-0.54^{***}
				(0.06)
R^2	0.74	0.74	0.76	0.78
Adj. R^2	0.73	0.73	0.75	0.78
Num. obs.	457	457	451	457

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Question 2 – (a)

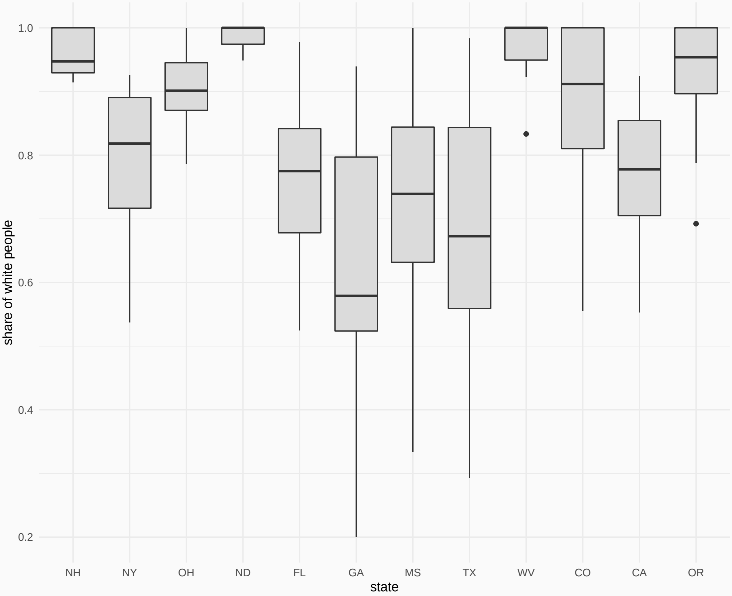
Panel data exercise. Provide summary statistics and plots of the 6 variables in the data set that are not the panel identifiers year and state. Plot `FTM` by state over time.

Question 2 – (a)

Summary statistics (pooled):

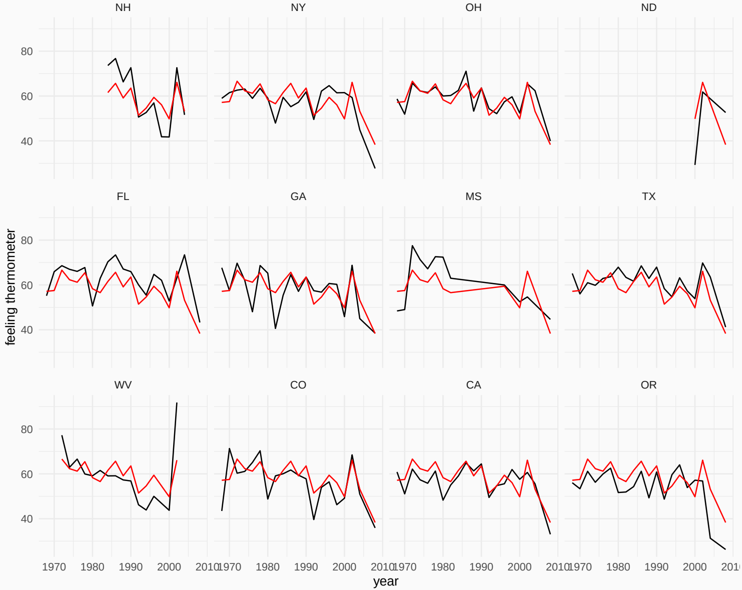
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
FTM	201	58.249	9.516	26.296	53.364	63.511	91.833
white	261	0.809	0.162	0	0.7	0.9	1
poor	261	0.181	0.117	0.000	0.109	0.227	0.692
turnout	260	1.654	0.166	1.000	1.559	1.779	2.000
voteDem	261	0.164	0.171	0.000	0.000	0.295	0.667
dem	261	0.282	0.195	0.000	0.000	0.417	0.737

Question 2 – (a)



1.0
0.8
0.6
0.4
0.2
share of white people
NH
NY
OH
ND
FL
GA
MS
TX
WV
CO
CA
OR
state

Question 2 – (a)


NH
NY
OH
ND
FL
GA
MS
TX
WV
CO
CA
OR
feeling thermometer
year
40
60
80
1970
1980
1990
2000
2010

Question 2 – (b)

Get overall, within, and between variation (standard deviation) of the variables in the data set from R or Stata.

Question 2 – (b)

Get overall, within, and between variation (standard deviation) of the variables in the data set from R or Stata.

```
overall.variation
    year      FTM     white      poor
15.66012 9.516032 0.1615002 0.1174961

within.variation
var.year  var.FTM var.white  var.poor
15.06844 9.070157 0.1257591 0.1014254

between.variation
year_g.mean FTM_g.mean white_g.mean poor_g.mean
7.81        3.98       0.115        0.0725
```

Question 2 – (c), (d), (e), (f), (g)

(c) Run a pooled OLS regression of `FTM` on `white`, `poor`, `dem`, and `turnout`. Explain why the estimated coefficients and standard errors may be biased.

(d) Re-run the regression above but allow errors to be clustered by `state`. How are the estimation results different? Explain why they are different.

(e) Re-run the regression above but include dummies for each state. How are the estimation results different? Explain why they are different.

(f) Re-run the regression above but use the fixed effects estimator (or within estimator).

(g) Re-run the regression above but use the random effects estimator.

Question 2 – (c), (d), (e), (f), (g)

	(c)	(d)	(e)	(f)	(g)
(Intercept)	83.91^{***}	83.91^{***}	50.52^{***}		58.84^{***}
	(8.24)	(11.07)	(7.86)		(5.87)
white	8.30^{**}	8.30^{*}	12.04^{*}	12.04^{*}	1.37
	(4.18)	(4.47)	(6.42)	(7.28)	(4.75)
poor	-5.66	-5.66	-1.83	-1.83	3.33
	(6.94)	(10.35)	(7.87)	(10.42)	(9.14)
dem	2.35	2.35	5.46	5.46	5.82^{*}
	(5.05)	(4.48)	(5.15)	(3.55)	(3.36)
turnout	-19.35^{***}	-19.35^{***}			
	(4.32)	(4.59)			
voteDem			-23.42^{***}	-23.42^{***}	-27.66^{***}
			(4.03)	(3.00)	(3.09)
R^2	0.10	0.10	0.29	0.22	0.22
Adj. R^2	0.08	0.08	0.23	0.16	0.20
Num. obs.	201	201	201	201	201

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Question 2 – (h)

Run the Hausman test and determine whether the fixed effects or the random effects model is more appropriate for the data at hand.

Question 2 – (h)

Run the Hausman test and determine whether the fixed effects or the random effects model is more appropriate for the data at hand.

```
     Hausman Test
data:  FTM ~ voteDem + dem + poor + white
chisq = 10.135, df = 4, p-value = 0.03822
alternative hypothesis: one model is inconsistent
```

- We reject the null, thus we should use a FE model.

Question 2 – (h)

Run the Hausman test and determine whether the fixed effects or the random effects model is more appropriate for the data at hand.

```
     Hausman Test
data:  FTM ~ voteDem + dem + poor + white
chisq = 10.135, df = 4, p-value = 0.03822
alternative hypothesis: one model is inconsistent
```

- We reject the null, thus we should use a FE model.
- **Yet** a Hausman test only cares about the precision of the estimate, and it is not everything we need to consider

Question 2 – (h)

Run the Hausman test and determine whether the fixed effects or the random effects model is more appropriate for the data at hand.

```
     Hausman Test
data:  FTM ~ voteDem + dem + poor + white
chisq = 10.135, df = 4, p-value = 0.03822
alternative hypothesis: one model is inconsistent
```

- We reject the null, thus we should use a FE model.
- **Yet** a Hausman test only cares about the precision of the estimate, and it is not everything we need to consider
- In this case we know we have lots of within and not so much between variation: a RE model still makes sense

Question 2 – (h)

Run the Hausman test and determine whether the fixed effects or the random effects model is more appropriate for the data at hand.

```
     Hausman Test
data:  FTM ~ voteDem + dem + poor + white
chisq = 10.135, df = 4, p-value = 0.03822
alternative hypothesis: one model is inconsistent
```

- We reject the null, thus we should use a FE model.
- **Yet** a Hausman test only cares about the precision of the estimate, and it is not everything we need to consider
- In this case we know we have lots of within and not so much between variation: a RE model still makes sense
- In these cases you can estimate both and hope to get similar results

Maximum Likelihood Estimation

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference
    - In probability theory we ask $p(y, \mathbf{X}|\theta)$

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference
   - In probability theory we ask $p(y, \mathbf{X}|\theta)$
   - Yet in reality we start from our data $(y, \mathbf{X})$ and we want to find out what the parameters are!

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference
    - In probability theory we ask $p(y, \mathbf{X}|\theta)$
    - Yet in reality we start from our data $(y, \mathbf{X})$ and we want to find out what the parameters are!
    - We want to find out the **likelihood** that the parameter is one that we estimate, given the data that we have.

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference
    - In probability theory we ask $p(y, \mathbf{X}|\theta)$
    - Yet in reality we start from our data $(y, \mathbf{X})$ and we want to find out what the parameters are!
    - We want to find out the **likelihood** that the parameter is one that we estimate, given the data that we have.
    - We are asking the same question, only starting from a different starting point: $p(y, \mathbf{X}|\theta) = \mathcal{L}(\theta|y, \mathbf{X})$

Introduction to MLE

What is MLE?

1. It's an estimation technique (we use it to estimate parameters, as in the LS framework)
2. It is a statistical theory of inference
    - In probability theory we ask $p(y, \mathbf{X}|\theta)$
    - Yet in reality we start from our data $(y, \mathbf{X})$ and we want to find out what the parameters are!
    - We want to find out the **likelihood** that the parameter is one that we estimate, given the data that we have.
    - We are asking the same question, only starting from a different starting point: $p(y, \mathbf{X}|\theta) = \mathcal{L}(\theta|y, \mathbf{X})$
    - To perform MLE you need to have some prior idea of what functional form the distribution of your dependent variable can follow

Four steps to MLE

1. Formulate the probability density function (PDF) of your y

Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i(\theta | y_i, \mathbf{X_i}) = \prod_{i=1}^{N} p_i(y_i, \mathbf{X_i} | \theta) \rightarrow log \mathcal{L} = \sum_{i=1}^{N} log[p_i(y_i, \mathbf{X_i} | \theta)]$$

Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i(\theta|y_i, \mathbf{X_i}) = \prod_{i=1}^{N} p_i(y_i, \mathbf{X_i}|\theta) \rightarrow log\mathcal{L} = \sum_{i=1}^{N} log[p_i(y_i, \mathbf{X_i}|\theta)]$$

3. Take the first derivative of $log\mathcal{L}$ and set to 0 (first order condition) to find a maximum

Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i(\theta | y_i, \mathbf{X_i}) = \prod_{i=1}^{N} p_i(y_i, \mathbf{X_i} | \theta) \rightarrow log\mathcal{L} = \sum_{i=1}^{N} log[p_i(y_i, \mathbf{X_i} | \theta)]$$

3. Take the first derivative of $log\mathcal{L}$ and set to 0 (first order condition) to find a maximum $\frac{\partial log\mathcal{L}}{\partial \theta} = 0$

Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i(\theta | y_i, \mathbf{X_i}) = \prod_{i=1}^{N} p_i(y_i, \mathbf{X_i} | \theta) \rightarrow log\mathcal{L} = \sum_{i=1}^{N} log[p_i(y_i, \mathbf{X_i} | \theta)]$$

3. Take the first derivative of $log\mathcal{L}$ and set to 0 (first order condition) to find a maximum $\frac{\partial log\mathcal{L}}{\partial \theta} = 0$
4. Take the second derivative and **verify** that it is smaller than 0 (second order condition) to verify it's a maximum (not a minimum)

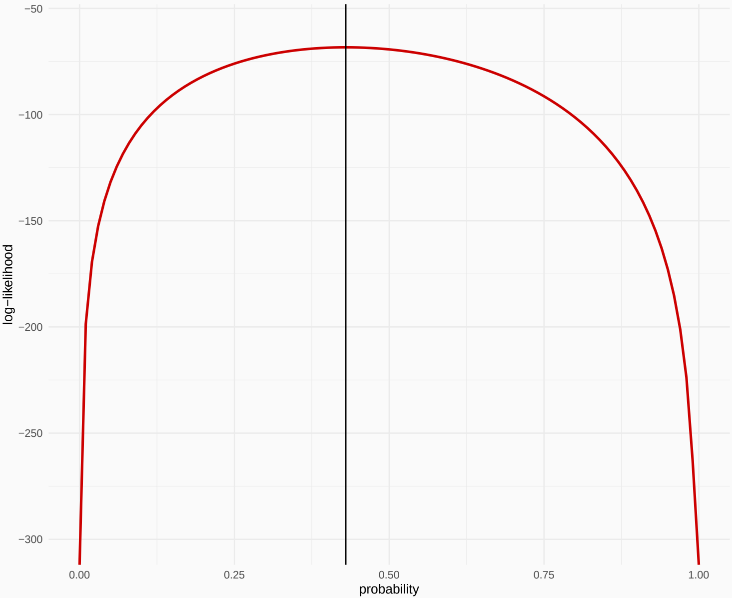
Four steps to MLE

1. Formulate the probability density function (PDF) of your y
2. Express the joint distribution of the observations with the PDF chosen and turn likelihoods into probabilities. Express the joint likelihood $\mathcal{L}$ as a log. (Assume i.i.d. observations):

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i(\theta|y_i, \mathbf{X_i}) = \prod_{i=1}^{N} p_i(y_i, \mathbf{X_i}|\theta) \rightarrow log\mathcal{L} = \sum_{i=1}^{N} log[p_i(y_i, \mathbf{X_i}|\theta)]$$

3. Take the first derivative of $log\mathcal{L}$ and set to 0 (first order condition) to find a maximum $\frac{\partial log\mathcal{L}}{\partial\theta} = 0$
4. Take the second derivative and **verify** that it is smaller than 0 (second order condition) to verify it's a maximum (not a minimum) $\frac{\partial^2 log\mathcal{L}}{\partial\theta} < 0$

Log-likelihood function of a binomial distribution



Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

$$\mathbf{H}(\theta) = \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\delta} \end{bmatrix}$$

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

$$\mathbf{H}(\theta) = \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\delta} \end{bmatrix}$$

- Interpret the diagonal as change in $log\mathcal{L}$ around α, β and δ

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

$$\mathbf{H}(\theta) = \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\delta} \end{bmatrix}$$

- Interpret the diagonal as change in $log\mathcal{L}$ around α, β and δ
- The lower the change (the curvature of the log-likelihood) the further you are from the optimum, the larger the variance

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

$$\mathbf{H}(\theta) = \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\delta} \end{bmatrix}$$

- Interpret the diagonal as change in $log\mathcal{L}$ around α, β and δ
- The lower the change (the curvature of the log-likelihood) the further you are from the optimum, the larger the variance
- The VCOV matrix of the MLE will be: $Var(\hat{\theta}) = -E[\mathbf{H}(\theta)]^{-1}$

Variance and standard error of the estimates

- How to get uncertainty of a θ vector of estimates (α, β, δ)?
- We use the Hessian matrix, a matrix of second derivatives:

$$\mathbf{H}(\theta) = \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\alpha\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\beta\partial\delta} \\ \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\alpha} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\beta} & \frac{\partial^2 log\mathcal{L}(\theta)}{\partial\delta\partial\delta} \end{bmatrix}$$

- Interpret the diagonal as change in $log\mathcal{L}$ around α, β and δ
- The lower the change (the curvature of the log-likelihood) the further you are from the optimum, the larger the variance
- The VCOV matrix of the MLE will be: $Var(\hat{\theta}) = -E[\mathbf{H}(\theta)]^{-1}$
- $SE(\hat{\theta})$ = squared root of the diagonal of the inverse of the negative of the Hessian matrix

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$f(y_i|\alpha + \beta x_i, \sigma) =$$

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$f(y_i|\alpha + \beta x_i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{[y_i - (\alpha + \beta x_i)]^2}{\sigma^2}\right)$$

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$
\begin{aligned}
f(y_i|\alpha + \beta x_i, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{[y_i - (\alpha + \beta x_i)]^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma}\phi\left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2}\right)
\end{aligned} \tag{1}
$$

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$
\begin{aligned}
f(y_i|\alpha+\beta x_i,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{[y_i-(\alpha+\beta x_i)]^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma}\phi\left(\frac{y_i-[\alpha+\beta x_i]}{\sigma^2}\right)
\end{aligned}
\tag{1}
$$

Where ϕ is the standard normal distribution:

$$
\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right)
$$

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$f(y_i|\alpha+\beta x_i,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{[y_i-(\alpha+\beta x_i)]^2}{\sigma^2}\right) \\ = \frac{1}{\sigma}\phi\left(\frac{y_i-[\alpha+\beta x_i]}{\sigma^2}\right) \tag{1}$$

Where ϕ is the standard normal distribution:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right)$$

Equation (1) gives the following log-likelihood function:

Re-estimating the linear model in a ML framework

The linear model can be re-estimated in a ML framework:

$$
\begin{aligned}
f(y_i|\alpha+\beta x_i,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{[y_i-(\alpha+\beta x_i)]^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma}\phi\left(\frac{y_i-[\alpha+\beta x_i]}{\sigma^2}\right)
\end{aligned} \tag{1}
$$

Where ϕ is the standard normal distribution:

$$
\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right)
$$

Equation (1) gives the following log-likelihood function:

$$
log\mathcal{L}(\alpha,\beta,\sigma|y_i,x_i) = \sum_{i=1}^{N} log\frac{1}{\sigma}\phi\left(\frac{y_i-[\alpha+\beta x_i]}{\sigma^2}\right)
$$

Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

$$log \mathcal{L}(\alpha, \beta, \sigma | y_i, x_i) = \sum_{i=1}^{N} log \frac{1}{\sigma} \phi \left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2} \right)$$

Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

$$log \mathcal{L}(\alpha, \beta, \sigma | y_i, x_i) = \sum_{i=1}^{N} log \frac{1}{\sigma} \phi \left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2} \right)$$

- Both R and Stata have functions that maximize a function, which we can use to "climb" the log-likelihood function until we reach a maximum (hopefully global, not local).

Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

$$log\mathcal{L}(\alpha, \beta, \sigma | y_i, x_i) = \sum_{i=1}^{N} log\frac{1}{\sigma}\phi\left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2}\right)$$

- Both R and Stata have functions that maximize a function, which we can use to "climb" the log-likelihood function until we reach a maximum (hopefully global, not local).
- We can follow a similar procedure and optimize whatever log-likelihood function we can design.

Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

$$log\,\mathcal{L}(\alpha, \beta, \sigma | y_i, x_i) = \sum_{i=1}^{N} log \frac{1}{\sigma} \phi \left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2} \right)$$

- Both R and Stata have functions that maximize a function, which we can use to "climb" the log-likelihood function until we reach a maximum (hopefully global, not local).
- We can follow a similar procedure and optimize whatever log-likelihood function we can design.
- Most often, though, we will use functions that rely on a particular log-likelihood specification (such as the functions of the GLM family).

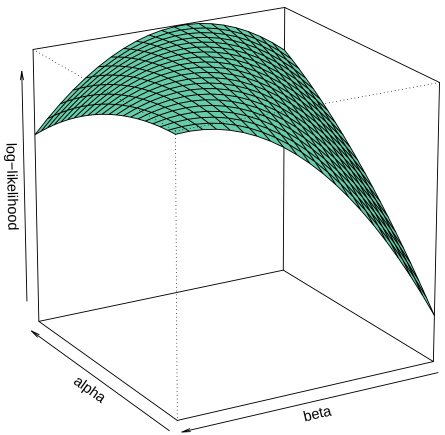
Optimize the linear log-likelihood function

- Using R or Stata we can maximize the linear log-likelihood function which we have obtained:

$$log\mathcal{L}(\alpha, \beta, \sigma | y_i, x_i) = \sum_{i=1}^{N} log\frac{1}{\sigma}\phi\left(\frac{y_i - [\alpha + \beta x_i]}{\sigma^2}\right)$$

- Both R and Stata have functions that maximize a function, which we can use to "climb" the log-likelihood function until we reach a maximum (hopefully global, not local).
- We can follow a similar procedure and optimize whatever log-likelihood function we can design.
- Most often, though, we will use functions that rely on a particular log-likelihood specification (such as the functions of the GLM family). We won't need to write a $log\mathcal{L}$ function

Log-likelihood function of a linear model with two parameters



Conclusion

All clear? More questions?

Thanks and see you next week!