

Gravity Lab:

(1) Intro to Stata.

Luca De Benedictis¹ and Luca Salvatici²

¹University of Macerata - debene@unimc.it

²University Roma Tre - luca.salvatici@uniroma3.it

Lecture 1 - Laboratory

Part I

Openness

Openness

It measures the degree of involvement in international real transaction of a country

$$O_{it} = \frac{X_{it} + M_{it}}{Y_{it}} \geq 0 \quad (1)$$

Font: Penn World Tables (PWTs)

Data: openness12.dta

Notes: the openness12.dta dataset includes openness $\times 100$.

Variables in *openness12.dta*

Stata:

```
. desc
```

```
Contains data from openness_12.dta
```

```
obs:      3,161
vars:      10                      2 Oct 2015 16:44
size:     148,567
```

variable name	storage type	display format	value label	variable label
ccode	str3	%9s		iso-code
year	int	%8.0g		year
openc	double	%10.0g		trade openness in current terms
openk	double	%10.0g		trade openness in real terms
pop	double	%9.0g		total population
gdp_current	double	%10.0g		gdp in current terms
ldlock	byte	%8.0g		1 if landlock
island	byte	%8.0g		1 if island
remoteness	float	%9.0g		remoteness
remoteness_head	float	%9.0g		remoteness according to Head

```
Sorted by:  ccode  year
```

Descriptive statistics

Stata:

```
. *Descriptive statistics with summarize  
. summarize openc gdp_current pop
```

Variable	Obs	Mean	Std. Dev.	Min	Max
openc	3148	69.55744	47.23225	4.83	428.44
gdp_current	3039	1.90e+11	7.57e+11	1.11e+08	1.17e+13
pop	3161	42947.93	137814.3	161.63	1300000

There are some missing observation. Where ?

Scatterplot: Openness and GDPpc

Stata:

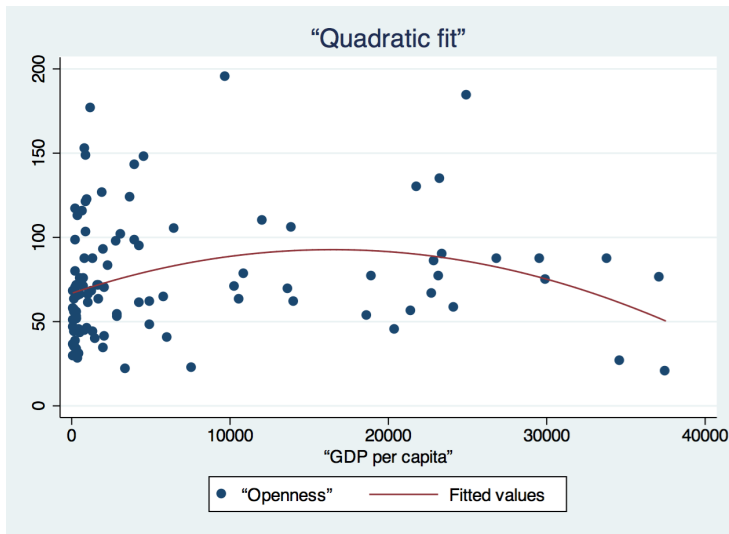
```
// use openness12.dta, replace
generate gdppc = gdp_current/pop
replace gdppc = gdppc/1000
generate ln_gdppc = ln(gdppc)

twoway (scatter openc gdppc) (qfit openc gdppc) if (year==2000 & openc<=200),
    title("Quadratic fit") legend(lab(1 "Openness")) xtitle ("GDP per capita")

twoway (scatter openc ln_gdppc) (qfit openc ln_gdppc) if (year==2000 & openc<=200),
    title("Quadratic fit after log transformation") legend(lab(1 "Openness"))
    xtitle ("log GDP per capita")
```

Openness: Stata 1

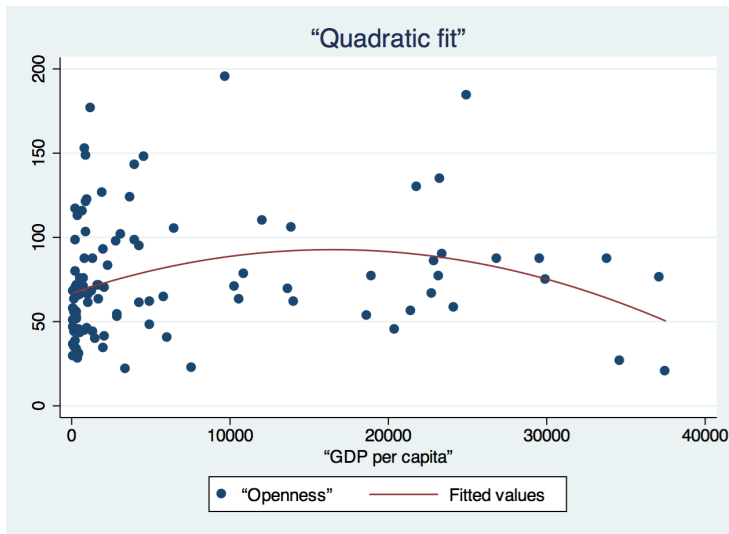
Openness vs GDP per capita



Why is not a good idea to fit a polynomial ?

Openness: Stata 1

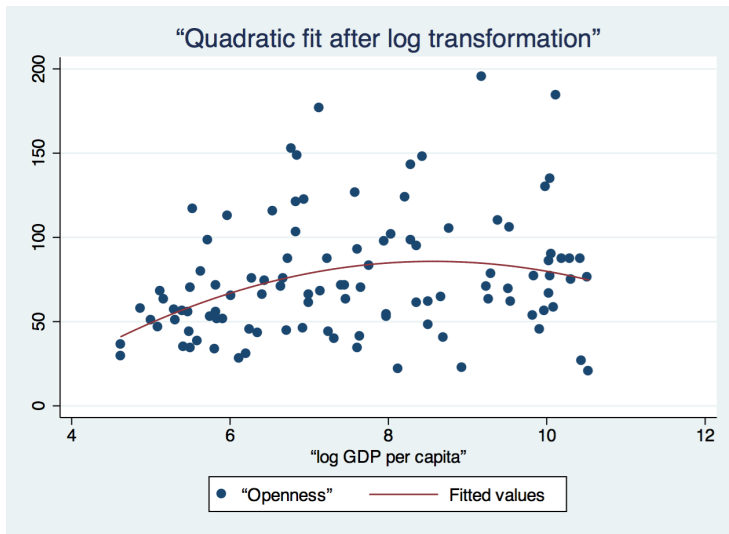
Openness vs GDP per capita



Why is not a good idea to fit a polynomial ?

Openness: Stata 2

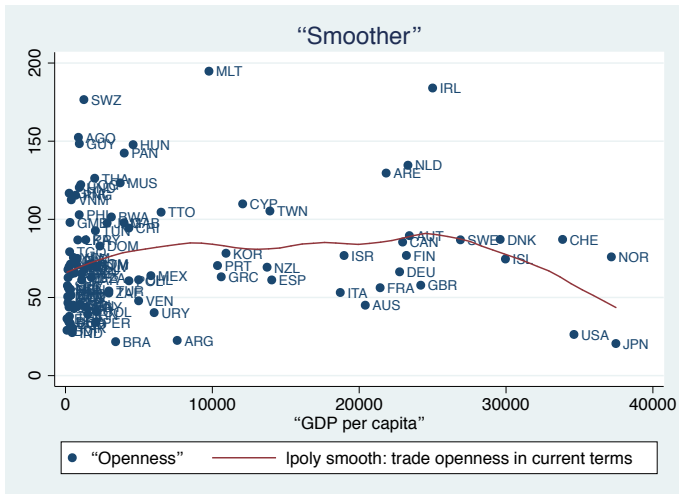
Openness vs GDP per capita (in log)



Why is not a good idea to fit a polynomial ?

Openness: Stata 3

Openness vs GDP per capita



Better to use a smoother!

Part II

Let's start with Stata