

Оценка роста СПФ на данных по российским предприятиям за 2008-2015 годы

Практикум по прикладным
экономическим исследованиям

Начало работы

1. Отрываем файл
`use "Z:\Data\tfp_1581.dta"`
2. Смотрим на базу
`d` (или `describe`)

Число наблюдений в базе

id – идентификационный код
предприятия
year – год
Каждое предприятие можем
наблюдать несколько лет

The screenshot shows the Stata/MP 14.2 interface. The command window displays the following commands and output:

```
. clear
. use "Z:\Data\tfp_1581.dta"
. d

Contains data from Z:\Data\tfp_1581.dta
obs:      8,266
vars:      6
size:     198,384
10 May 2019 22:04
```

variable name	storage type	display format	value label	variable label
id	float	%9.0g		Firm ID
year	float	%ty		
ind4d	float	%9.0g		OKVED, 4 digit
rva	float	%9.0g		Real Value Added
emp	long	%12.0g		Number of Employees
rk	float	%9.0g		Real Capital

Sorted by:

Command

Variables

Name	Label
id	Firm ID
year	
ind4d	OKVED, 4 digit
rva	Real Value Added
emp	Number of Emplo...
rk	Real Capital

Properties

Variables

Name	Label

Data

CAP NUM OVR

Z:\Data

Переменные и их описание

Начало работы

Смотрим на данные

sort id year (упорядочиваем данные по предприятию и году)

edit

The screenshot displays the Stata 14.2 interface. The main window shows the 'Data Editor (Edit)' for the file 'tftp_1581.dta'. The data is sorted by 'id' and 'year'. The command window on the left shows the following commands:

```
clear
use "Z:\Data\tftp_1581.dta"
d
sort id year
edit
```

The command window also displays the following information:

```
Contains data from Z:\Data\tftp_1581.dta
obs:      8,266
vars:      6
size:     198,384
```

The variable list shows the following variables and their types:

variable name	type
id	float
year	float
ind4d	float
rva	float
emp	long
rk	float

The 'Data Editor' window shows the following data:

	id	year	ind4d	rva	emp	rk
1	1904	2014	1581	590431.5	17	1026872
2	1904	2015	1581	88803.2	22	892022.1
3	1933	2008	1581	7579525	29	897824
4	1933	2009	1581	1.05e+07	36	1037543
5	1933	2010	1581	1.18e+07	43	1194253
6	1933	2011	1581	1.24e+07	50	1645218
7	1933	2012	1581	1.10e+07	55	1295347
8	1933	2013	1581	1.31e+07	66	2253729
9	1936	2008	1581	2.56e+07	155	6576159
10	1936	2010	1581	1638948	167	1401654
11	2704	2014	1581	1.40e+07	515	6.02e+07
12	2704	2015	1581	1.48e+07	498	9.19e+07
13	2711	2008	1581	3.10e+07	317	8.10e+07
14	2711	2010	1581	4.31e+07	321	5.54e+07
15	2711	2011	1581	3.14e+07	310	6.02e+07
16	2711	2012	1581	2.42e+07	299	4.66e+07
17	2711	2013	1581	1.99e+07	288	3.75e+07
18	2766	2008	1581	1.31e+07	57	8751182
19	2766	2009	1581	1.41e+07	62	1.04e+07
20	2766	2010	1581	1.58e+07	142	1.25e+07
21	2766	2011	1581	1.89e+07	167	1.21e+07
22	2766	2012	1581	1.90e+07	176	1.08e+07
23	2766	2013	1581	3.37e+07	187	1.53e+07
24	2766	2015	1581	2.30e+07	183	1.45e+07
25	2767	2011	1581	6460991	11	3025363
26	2767	2012	1581	1.37e+07	17	4450463
27	2767	2013	1581	1.05e+07	14	1.05e+07
28	2767	2014	1581	9763870	19	1.05e+07
29	2767	2015	1581	9260464	21	1.05e+07
30	2768	2008	1581	1.40e+07	285	1.40e+07

The 'Variables' window on the right shows the following variables and their labels:

Name	Label
id	Firm ID
year	
ind4d	OKVED, 4 digit
rva	Real Value Added
emp	Number of Emplo...
rk	Real Capital

The 'Properties' window on the right shows the following properties for the variable 'id':

Property	Value
Name	id
Label	Firm ID
Type	float
Format	%9.0g
Value label	
Notes	

The 'Data' window on the right shows the following data:

Variable	Value
id	1904
year	2014
ind4d	1581
rva	590431.5
emp	17
rk	1026872

Red arrows point from the text annotations to the corresponding data rows in the 'Data Editor' window.

Данные для фирмы есть только в конце периода

Данные для фирмы есть только за 2008 и 2010 годы

Данные для фирмы есть почти за все годы с 2008 по 2015 (кроме 2014)

Описание данных

Предприятия в отрасли «Производство хлеба и мучных кондитерских изделий недлительного хранения» (код ОКВЭД 1581)

Период наблюдения 2008-2015 годы

Не все предприятия мы наблюдаем в каждом году – несбалансированная панель

. tab year

year	Freq.	Percent	Cum.
2008	879	10.63	10.63
2009	913	11.05	21.68
2010	881	10.66	32.34
2011	919	11.12	43.46
2012	964	11.66	55.12
2013	1,164	14.08	69.20
2014	1,273	15.40	84.60
2015	1,273	15.40	100.00
Total	8,266	100.00	

Переменные:

rva	Real Value Added	Реальная добавленная стоимость, рассчитана на основе данных бухгалтерских балансов предприятия, показатель дефлирован с помощью индекса цен производителей (ИПЦ)
emp	Number of Employees	Число работников на предприятии
rk	Real Capital	Реальная стоимость основных фондов, данные бухгалтерских балансов предприятия, дефлированы на основе индексов физического объема основных фондов

Панельные данные

Задаем панельную структуру данных

Пространственная переменная – предприятие (id)

Временная переменная – год (year)

xtset id year, yearly

Периодичность наблюдения –
годовые данные

Stata/MP 14.2 - Z:\Data\tfp_1581.dta

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

#	Command	_rc
2	clear	
3	use "Z:\Data\tfp_1581.dta"	
4	d	
5	sort id year	
6	edit	
7	tab year	
8	xtset id year, yearly	

Sorted by:

```
. sort id year
```

```
. edit
```

```
. tab year
```

year	Freq.	Percent	Cum.
2008	879	10.63	10.63
2009	913	11.05	21.68
2010	881	10.66	32.34
2011	919	11.12	43.46
2012	964	11.66	55.12
2013	1,164	14.08	69.20
2014	1,273	15.40	84.60
2015	1,273	15.40	100.00
Total	8,266	100.00	

```
. xtset id year, yearly
```

```
panel variable: id (unbalanced)
```

```
time variable: year, 2008 to 2015, but with gaps
```

```
delta: 1 year
```

Variables

Name	Label
id	Firm ID
year	
ind4d	OKVED, 4 digit
rva	Real Value Added
emp	Number of Emplo...
rk	Real Capital

Properties

Variables

Name	Label
Type	
Format	
Value label	
Notes	

Data

CAP NUM OVR

Создаем новые переменные для оценки производственных функций

Для оценки функции Кобба-Дугласа

gen lva=ln(rva) логарифм реальной добавленной

gen lk=ln(rk) логарифм капитала

gen ll=ln(emp) логарифм труда

Для оценки транслогарифмической функции

gen lksq=lk^2 логарифм капитала в квадрате

gen llsq=ll^2 логарифм труда в квадрате

gen ll_lk=ll*lk произведение логарифма труда на логарифм капитала

Оценка регрессий на панельных данных

Оценка методом постоянных эффектов (**Fixed Effects Estimations**)

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + \alpha_i + u_{it}, t = 1, 2, \dots, T \quad (1)$$

Усредним это уравнение для каждой фирмы i за все годы

$$\bar{y}_i = \beta_1 \bar{x}_{i1} + \beta_2 \bar{x}_{i2} + \dots + \beta_k \bar{x}_{ik} + \alpha_i + \bar{u}_i \quad (2)$$

где $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$ и т.д.

В уравнение (2) входит α_i , так как этот параметр не изменяется во времени

Вычитаем из уравнения (1) уравнение (2)

$$y_{it} - \bar{y}_i = \beta_1 (x_{it1} - \bar{x}_{i1}) + \beta_2 (x_{it2} - \bar{x}_{i2}) + \dots + \beta_k (x_{itk} - \bar{x}_{ik}) + (u_{it} - \bar{u}_i)$$

или

$$\ddot{y}_i = \beta_1 \ddot{x}_{i1} + \beta_2 \ddot{x}_{i2} + \dots + \beta_k \ddot{x}_{ik} + \ddot{u}_i \quad (3)$$

Не содержит α_i можно оценить с помощью pooled OLS

где $\ddot{y}_i = y_{it} - \bar{y}_i$ и т.д. (time-demeaned data or **within transformation**)

Fixed effects estimation:

Оценка уравнения (3) с преобразованными данными методом pooled OLS

Команда в Стате для оценки методом постоянных эффектов

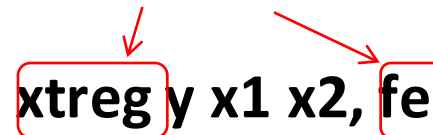
Мы уже задали панельную структуру данных (xtset...).

Можем использовать команды, которые работают с панельными данными

Fixed effects estimations

Оценка регрессии методом постоянных эффектов

xtreg y x1 x2, fe



(данные не надо трансформировать вручную, эта операция встроена в команду)

Оценка функции Кобба-Дугласа

$$\ln RVA_{it} = \alpha_0 + \alpha_L \ln L_{it} + \alpha_K \ln RK_{it}$$

методом постоянных эффектов

```
. xtreg lva ll lk, fe
```

```
Fixed-effects (within) regression
Group variable: id
```

```
R-sq:
  within = 0.1160
  between = 0.4746
  overall = 0.4756
```

```
corr(u_i, Xb) = 0.2312
```

```
Number of obs   = 8,266
Number of groups = 2,236
```

```
Obs per group:
   min = 1
   avg = 3.7
   max = 8
```

```
F(2,6028) = 395.65
Prob > F   = 0.0000
```

	lva	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	ll	.6162734	.0250255	24.63	0.000	.5672145	.6653324
	lk	.0924396	.0096597	9.57	0.000	.0735032	.111376
	_cons	11.23137	.1580284	71.07	0.000	10.92158	11.54117
	sigma_u	1.0057053					
	sigma_e	.4878627					
	rho	.80950864	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(2235, 6028) = 12.30      Prob > F = 0.0000
```

Число наблюдений в выборке.

Одну фирму мы наблюдаем несколько лет

Число фирм в выборке

Некоторые фирмы мы наблюдаем только один год
Часть фирм мы наблюдаем все восемь лет
В среднем для одной фирм мы располагаем данными за 3-4 года

Эластичность выпуска по труду

Эластичность выпуска по капиталу

xtreg lva ll lk, fe

Оценка транслогарифмической производственной функции методом постоянных эффектов

```
. xtreg lva ll lk llsq lksq ll_lk, fe
```

Fixed-effects (within) regression

Group variable: id

Number of obs = 8,266

Number of groups = 2,236

R-sq:

within = 0.1262

between = 0.4449

overall = 0.4532

Obs per group:

min = 1

avg = 3.7

max = 8

F(5,6025) = 173.96

Prob > F = 0.0000

corr(u_i, Xb) = -0.0021

lva	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ll	-.2102225	.1685363	-1.25	0.212	-.540614	.120169
lk	-.2916114	.0749734	-3.89	0.000	-.4385861	-.1446367
llsq	.1252894	.0210086	5.96	0.000	.0841051	.1664737
lksq	.0147011	.0029662	4.96	0.000	.0088863	.020516
ll_lk	-.0049141	.0103085	-0.48	0.634	-.0251224	.0152942
_cons	15.13288	.6057081	24.98	0.000	13.94548	16.32029
sigma_u	1.0065541					
sigma_e	.48518464					
rho	.81145877	(fraction of variance due to u_i)				

F test that all u_i=0: F(2235, 6025) = 12.21

Prob > F = 0.0000

Расчет роста СПФ на основе оценок транслогарифмической функции

- Рассчитываем эластичности выпуска по труду и капиталу (они будут отличаться для каждого предприятия и в разные годы)

```
gen eyk=_b[lk]+2*_b[lksq]*lk+_b[lk_lk]*lk
gen eyl=_b[l1]+2*_b[l1sq]*l1+_b[l1_lk]*lk
```

Обращение к коэффициенту перед переменной `l1_lk` в последней оцененной регрессии

- Рассчитываем прирост выпуска и факторов производства как разность логарифмов за два последовательных периода

```
gen dva=lva-l.lva
gen dk=lk-l.lk
gen dl=l1-l.l1
```

`l.` оператор лага, рассчитывает значение переменной в предыдущем периоде (если задана панельная структура данных или временные ряды)

- Рассчитываем среднюю эластичность факторов за два года

```
gen avg_eyk=(eyk+l.eyk)/2
gen avg_eyl=(eyl+l.eyl)/2
```

T-test на равенство 1 средней отдачи от масштаба

Отдача от масштаба: $E_{YKi} + E_{YLi}$

gen rts=eyp+eyl

```
. ttest rts==1 if year==2009
```

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
rts	913	.8408816	.0090697	.2740478	.8230818	.8586814

```
mean = mean(rts)                                t = -17.5440
Ho: mean = 1                                     degrees of freedom = 912
```

Ha: mean < 1	Ha: mean != 1	Ha: mean > 1
Pr(T < t) = 0.0000	Pr(T > t) = 0.0000	Pr(T > t) = 1.0000

Интерпретация результатов теста

<https://stats.idre.ucla.edu/stata/output/t-test/>

Расчет роста СПФ на основе оценок транслогарифмической функции

- Рассчитываем рост СПФ

```
gen tfp=dva-avg_eyk*dk-avg_eyl*d1
```

- Строим таблицы со средними и средневзвешенными (по реальной добавленной стоимости) показателями роста СПФ

```
tabstat tfp, by(year) statistics(n mean)
```

```
tabstat tfp [w=rva], by(year) statistics(n mean)
```

В квадратных скобках задаем веса, равные реальной добавленной стоимости

```
. gen tfp=dva-avg_eyk*dk-avg_eyl*d1
(2,637 missing values generated)

. tabstat tfp, by(year) statistics(n mean)
```

Summary for variables: tfp
by categories of: year

year	N	mean
2008	0	.
2009	736	.0238772
2010	762	-.165858
2011	727	-.1161348
2012	638	-.1094116
2013	843	-.0339785
2014	939	-.1024379
2015	984	-.0949763
Total	5629	-.0855098

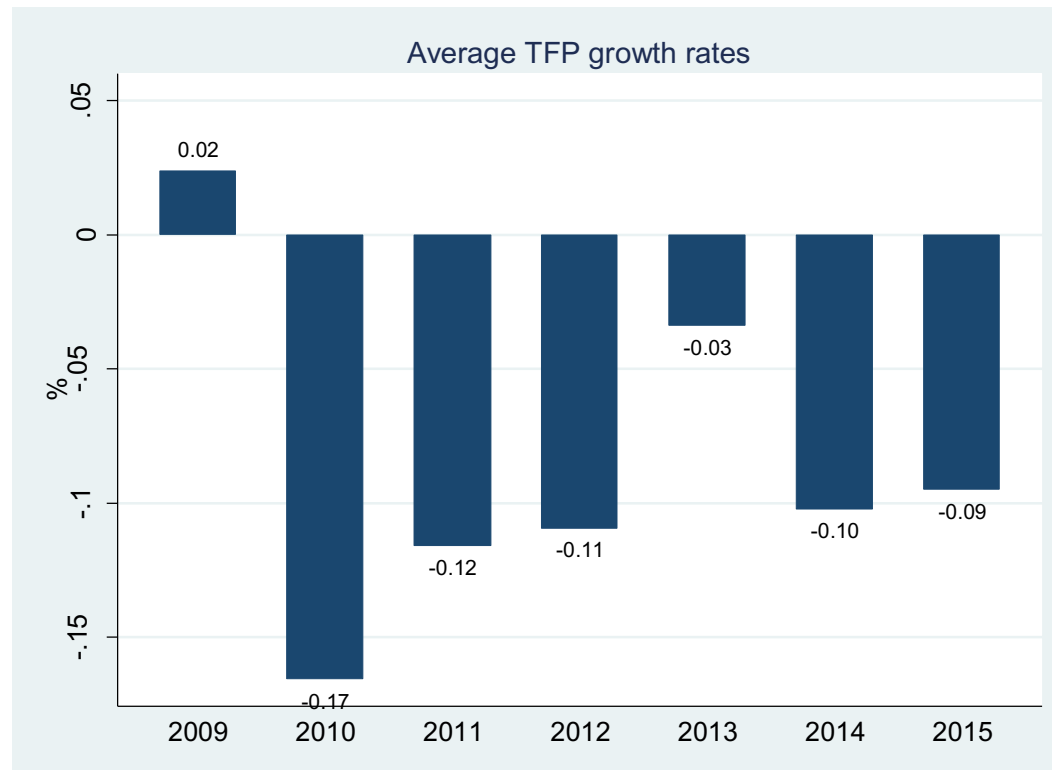
```
. tabstat tfp [w=rva], by(year) statistics(n mean)
(analytic weights assumed)
```

Summary for variables: tfp
by categories of: year

year	N	mean
2008	0	.
2009	736	.114145
2010	762	-.076123
2011	727	-.0485885
2012	638	-.0194532
2013	843	.0486897
2014	939	-.0051986
2015	984	.0173127
Total	5629	.0047964

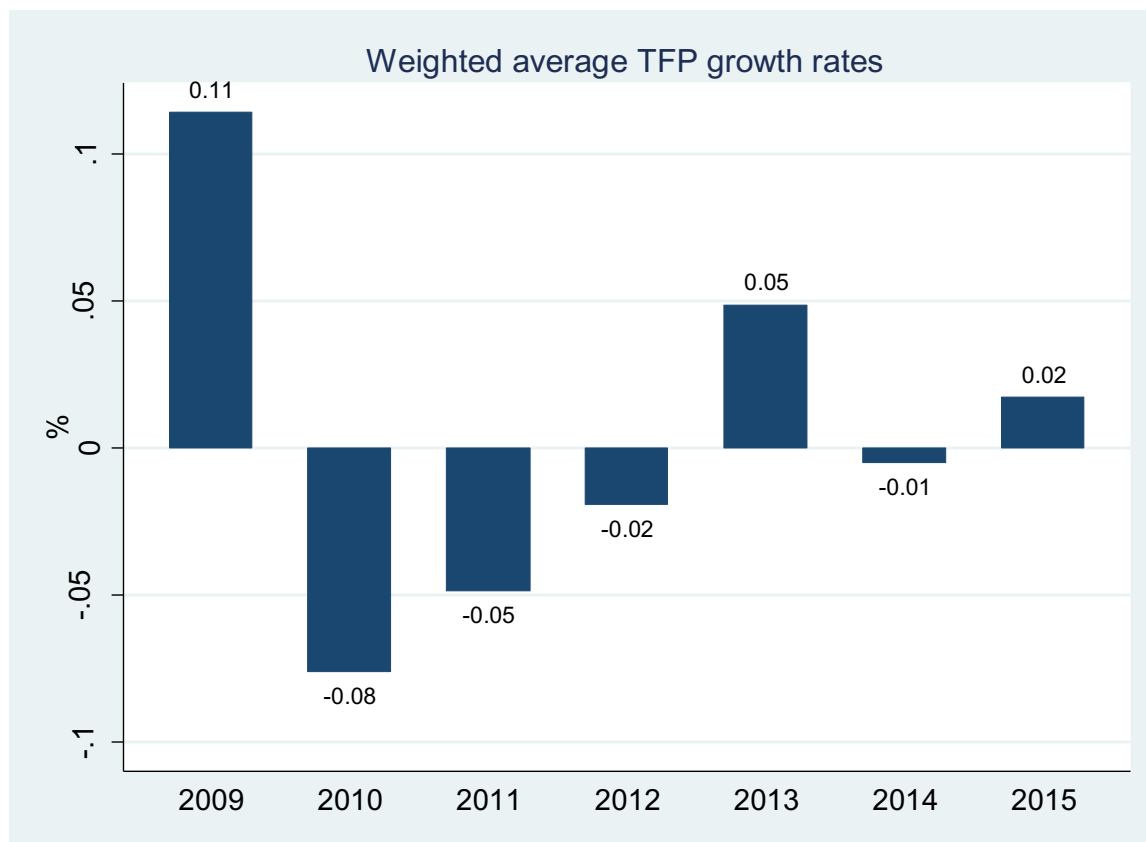
Средние темпы роста СПФ

```
graph bar tfp if year>=2009, over(year) name(gr1, replace) ///  
ytitle("%") blabel (bar, position(outside) format(%9.2f)) ///  
title("Average TFP growth rates", size(medium))
```



Средневзвешенные темпы роста СПФ

```
graph bar tfp [w=rva] if year>=2009, over(year) name(gr2, replace) ///  
ytitle("%") xlabel (bar, position(outside) format(%9.2f)) ///  
title("Weighted average TFP growth rates", size(medium))
```



Уровень эффективности на основе показателей производительности труда

Рассчитываем уровень производительности труда

gen lab_prod= rva/emp

Рассчитываем максимальный уровень производительности труда в каждом году

egen max_lab_prod=max(lab_prod), by(year)

Для каждого предприятия рассчитываем расстояние до максимального уровня
производительности труда (как долю от максимально возможной производительности труда)

gen lab_prod_gap=lab_prod/max_lab_prod

Смотрим на показатель эффективности

tabstat lab_prod_gap, by(year) statistics(n mean sd median min max p75)

```
. gen lab_prod= rva/emp
. egen max_lab_prod=max(lab_prod), by( year)
. gen lab_prod_gap=lab_prod/max_lab_prod
.
. tabstat lab_prod_gap, by(year) statistics(n mean sd median min max p90)
```

Summary for variables: lab_prod_gap
by categories of: year

year	N	mean	sd	p50	min	max	p90
2008	879	.1743485	.1518592	.1324489	.0076369	1	.3679195
2009	913	.1888542	.1647401	.141111	.0095084	1	.4072718
2010	881	.1590046	.1544964	.1118689	.0087054	1	.3399792
2011	919	.1319999	.1459403	.0838825	.0062685	1	.3027978
2012	964	.1177972	.1432013	.0693922	.0023697	1	.2799171
2013	1164	.1440246	.1643602	.0847833	.0032521	1	.3554963
2014	1273	.1568662	.1766316	.0907539	.0047164	1	.4042912
2015	1273	.1695016	.1830146	.1042093	.0053999	1	.412409
Total	8266	.155303	.1641242	.0993511	.0023697	1	.3614495

Рост СПФ по группам эффективности

- Создаем бинарную переменную, которая равна 1 для 10% наиболее эффективных предприятия и 0 в противном случае

(в **каждом году** находим уровень эффективности, который соответствует предприятию ниже которого находится **90% выборки**; создаем бинарную переменную равную 1 в случае, если уровень эффективности предприятия **выше**, чем заданная величина)

```
egen lpg90= pctlile(lab_prod_gap), p(90) by(year)
```

```
gen dummy_lpg90=(lab_prod_gap>=lpg90) if lab_prod_gap!=.
```

- Смотрим на рост СПФ по двум группам

1) Наиболее эффективные предприятия (dummy_lpg90==1)

2) Все остальные предприятия(dummy_lpg90==0)

```
tabstat tfp if dummy_lpg90==1, by(year) statistics(mean n)
```

```
tabstat tfp if dummy_lpg90==0, by(year) statistics(mean n)
```

```
. tabstat tfp if dummy_lpg90==1, by(year) statistics(mean n)
```

Summary for variables: tfp
by categories of: year

year	mean	N
2008	.	0
2009	.1731779	69
2010	.0064041	71
2011	-.0023985	54
2012	.1544003	30
2013	.1196491	60
2014	.0091167	87
2015	.0810004	98
Total	.0699714	469

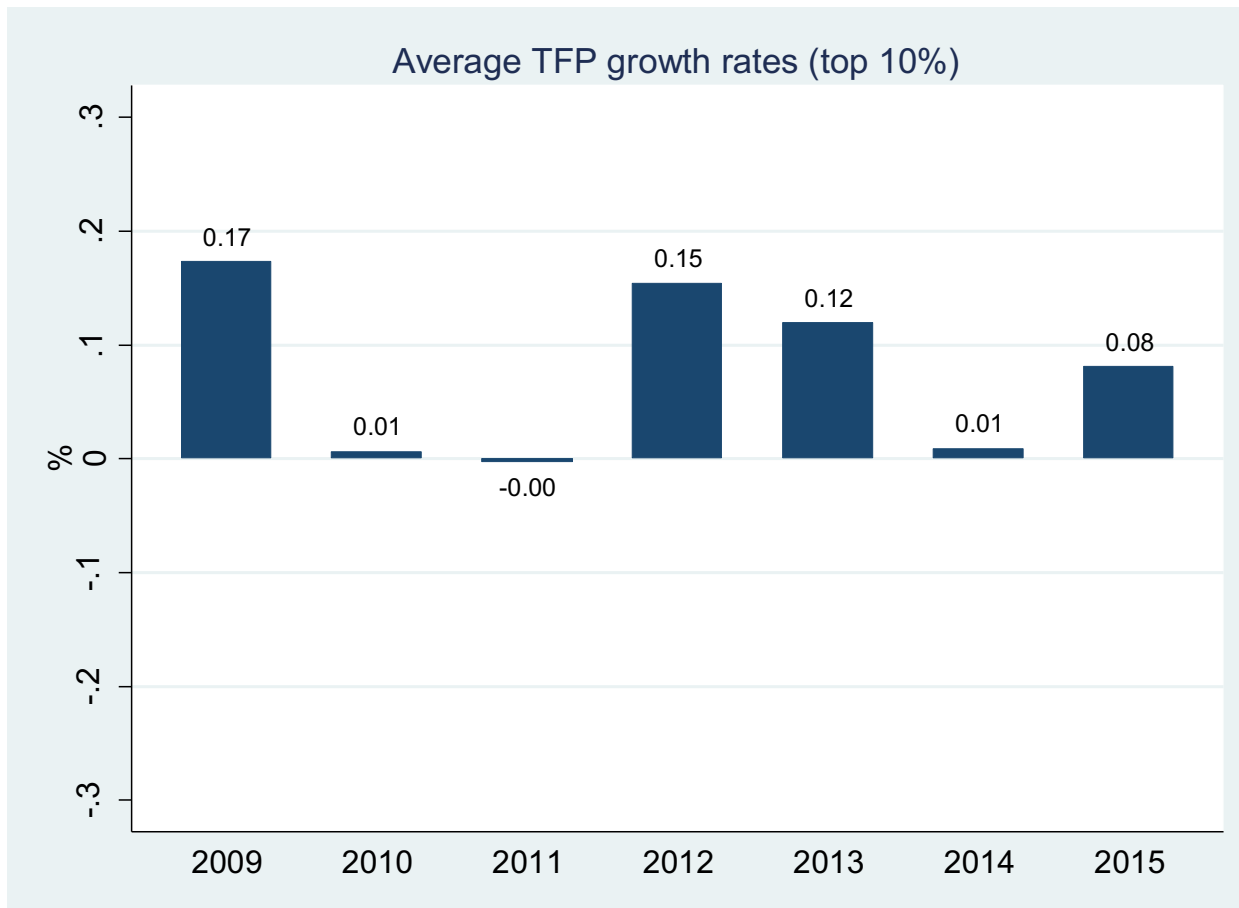
```
. tabstat tfp if dummy_lpg90==0, by(year) statistics(mean n)
```

Summary for variables: tfp
by categories of: year

year	mean	N
2008	.	0
2009	.0084323	667
2010	-.1835578	691
2011	-.1252607	673
2012	-.1224287	608
2013	-.0457507	783
2014	-.1138291	852
2015	-.114441	886
Total	-.0996417	5160

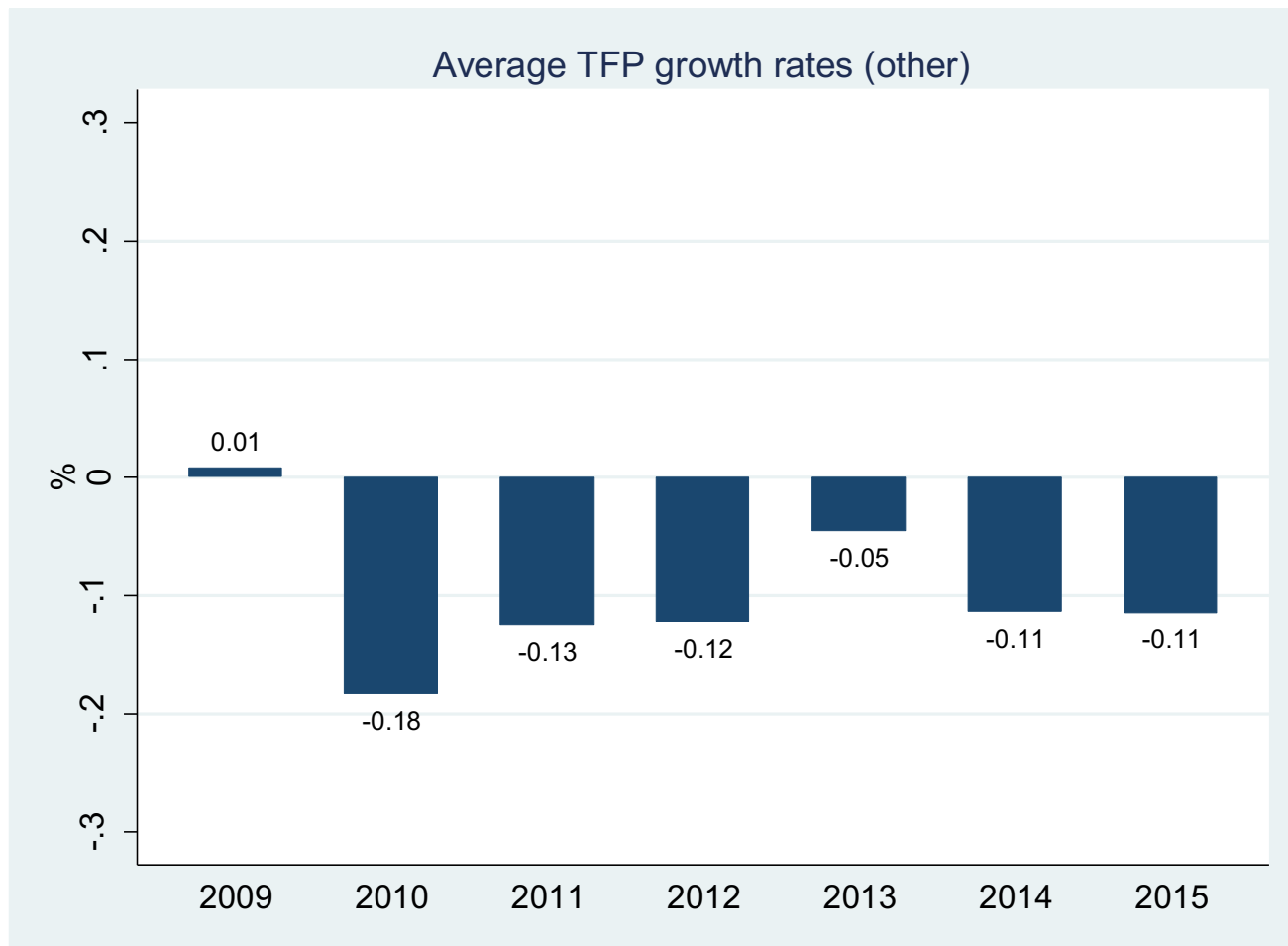
Средние темпы роста: наиболее эффективные предприятия

```
graph bar tfp if year>=2009 & dummy_lpg90==1, over(year) ///  
ytitle("%") blabel (bar, position(outside) format(%9.2f)) ylabel(-0.3(0.1)0.3) ///  
title("Average TFP growth rates (top 10%)", size(medium))
```

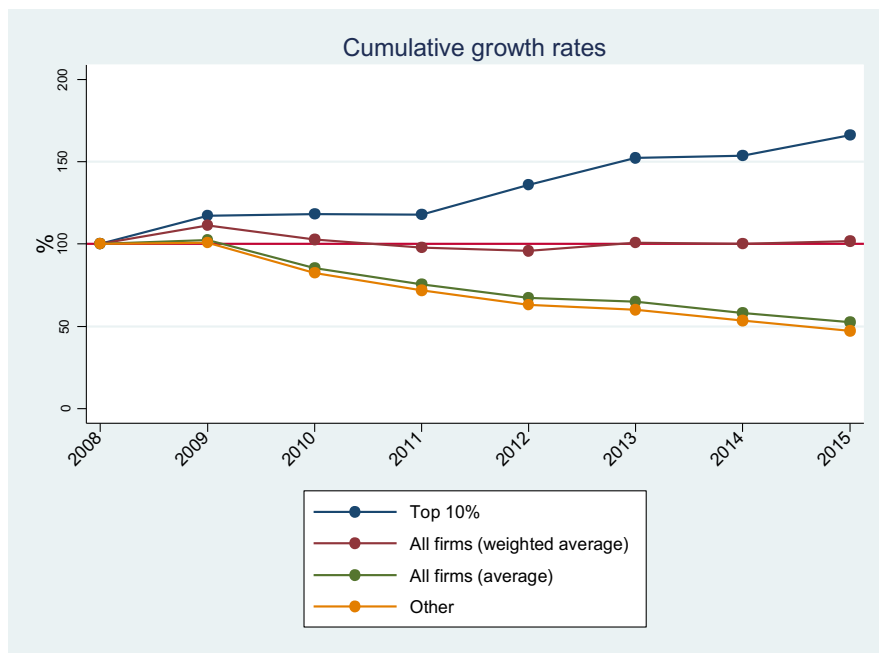


Средние темпы роста: прочие предприятия

```
graph bar tfp if year>=2009 & dummy_lpg90==0, over(year) ///  
ytitle("%") blabel (bar, position(outside) format(%9.2f)) ylabel(-0.3(0.1)0.3) ///  
title("Average TFP growth rates (other)", size(medium))
```



Накопленные темпы роста по группам эффективности



1. Наиболее эффективные предприятия показывают положительные темпы роста во все годы
2. Прочие фирмы начиная с 2010 года показывают отрицательные темпы роста
3. Увеличивается разрыв между наиболее эффективными предприятиями и всеми остальными
4. Неэффективные предприятия теряют долю на рынке. При этом они не выходят с рынка, продолжая неэффективно использовать факторы производства. В результате средние темпы роста в отрасли оказываются отрицательными, несмотря на высокие темпы роста у лидеров.