

29. новембар 2016. године

Оцена мера аграрне политике -
Статистичке основе анализе и примена модела

Теме:

Подаци панела

- Дефиниција
- Предности / ограничења података панела
- Основни модели
 - Модел са константним регресионим параметрима (енг. *Pooled model*)
 - Модел фиксних ефеката (енг. *Fixed effects model*)
- Примена у анализи економског положаја пољопривредних газдинстава у Србији



Врсте статистичких серија података

- (1) Упоредни подаци, структурне серије или подаци пресека (енг. *Cross – section data*)
- (2) Временске серије (енг. *Time - series*)
- (3) Подаци панела (енг. *Panel data*) – комбинација (1) и (2) (енг. *Pooling of cross sections over time*)

Шта су подаци панела?

Подаци панела (енг. *panel data; longitudinal data*):

- Комбинација упоредних података и временских серија
- Променљива X_{it} варира по две димензије – просторној и временској:

$$i=1,2,\dots,N; \quad t=1,2,\dots,T$$

- Подаци великог броја *истих* јединица посматрања у сукцесивним временским периодима
- Поновљене анкете једног истог узорка током времена.



У којим областима се користе подаци панела?

Микроекономска анализа (јединице посматрања: појединачна предузећа, банке, појединачни потрошачи, домаћинства, запослени, пољопривредна газдинства и сл.).

Макроекономска анализа (јединице посматрања: земље, региони, привредне гране, трговински сектори...).

Врсте података панела

- Класични ($N \gg T$) - “*short panels*”
- Панели са малим бројем јединица посматрања у дугом временском периоду (N и T велико или $T \gg N$) – “*large panels*”
- Балансирани и небалансирани панели
- Ротациони панели (*Spanish household survey: Spanish income/savings study*) и псеудо панели

Пример података панела

1. Балансирани и небалансирани панели

Земља	Ред.бр.	Година	Инфлација (%)	Стопа незапослености младих*
Белгија	1	2012	2.8	19.8
	1	2013	1.1	23.7
	1	2014	0.3	23.2
	1	2015	0.6	...
Македонија	2	2012	3.3	53.9
	2	2013	2.8	51.9
	2	2014	-0.3	53.1
	2	2015	-0.3	...
Србија	3	2012	7.3	51.1
	3	2013	7.7	49.4
	3	2014	2.1	...
	3	2015	1.4	...

* % укупне радне снаге 15-24 год.

Извор: World Development Indicators

Предности у односу на структурне и временске серије

- Истовремена анализа :
 - структуре и промена у структури током времена (индивидуални и временски ефекти).
 - ефеката фактора који не могу да се идентификују у случају модела временских серија и упоредних података.
- дужина временске серија недовољна или када је број јединица посматрања мали;
- повећава се узорак, расте бр. степени слободе, максимум информација;
- већи варијабилитет и ефикасност (прецизност) оцена, ублажавају се економетријски проблеми (нпр. мултиколинеарност, хетероскедастичност...);
- не постоји проблем избора репрезентативног периода.



Ограничења/проблеми у коришћењу података панела

- Проблем расположивости података панела, нарочито на микро нивоу (потрошачи, предузећа, домаћинства,...) - недостајући подаци у појединим периодима
- Могућности превазилажења проблема: коришћење других врста панела (нпр. небалансирани, ротациони и сл.)
- Сложенији економетријски проблеми:
 - проблеми својствени и упоредним подацима и временским серијама.

Линеарни модели панела

$$y_{it} = \beta_{1it} + \sum_{k=2}^K \beta_{kit} x_{kit} + u_{it}$$

Сви регр. параметри
константни (*pooled model*)

Регр. параметри варијабилни
по јединицама и/или кроз
време (фискни или стохаст.)

Слободни члан варира,
Константни параметри уз
регресоре

$$\beta_{kit} = \beta_k \quad \text{за } k=2, \dots, K$$

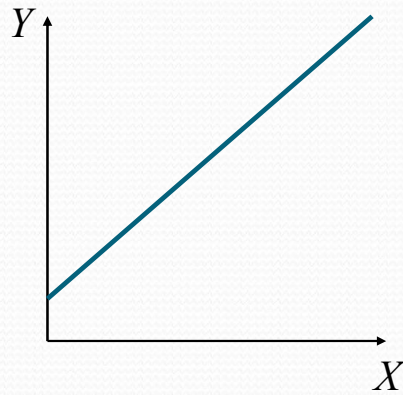
Модели индивидуалних
ефеката

Модели индивидуалних
и временских ефеката

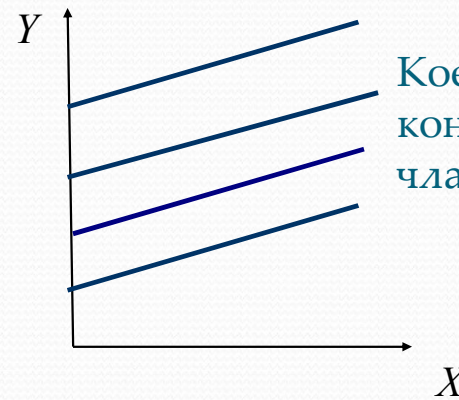
- Слободни члан варира само по јединицама посматрања

- Слободни члан варира само по јединицама посматрања и кроз време

Ограничења на параметре у моделу панела



Сви регр.
параметри
константни
(претпоставка
ретко испуњена у
пракси)



Коефицијент нагиба
константан, слободни
члан варијабилан

Модели панела

1. Модел са константним регресионим параметрима
(*pooled model*)

$$y_{it} = \beta_1 + \sum_{k=2}^K \beta_k x_{kit} + u_{it}$$

2. Модели са варијабилним слободним члановима (параметри уз регресоре константни) :

- Модели индивидуалних ефеката

$$y_{it} = \beta_{1i} + \sum_{k=2}^K \beta_k x_{kit} + u_{it}$$

- Модели индивидуалних и временских ефеката

$$y_{it} = \beta_{1it} + \sum_{k=2}^K \beta_k x_{kit} + u_{it}$$

Модел са константним регресионим параметрима - пример 1-

• На основу података о четири компаније: General Electric (GE), General Motors (GM), U.S. Steel (US), Westinghouse (West) за период од 20 година, оцењује се следећа једначина:

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + u_{it}$$

Зависна варијабла: Бруто инвестиције, I

Регресори: Тржишна вредност компаније, F
Вредност фиксних средстава (капиталне опреме), C

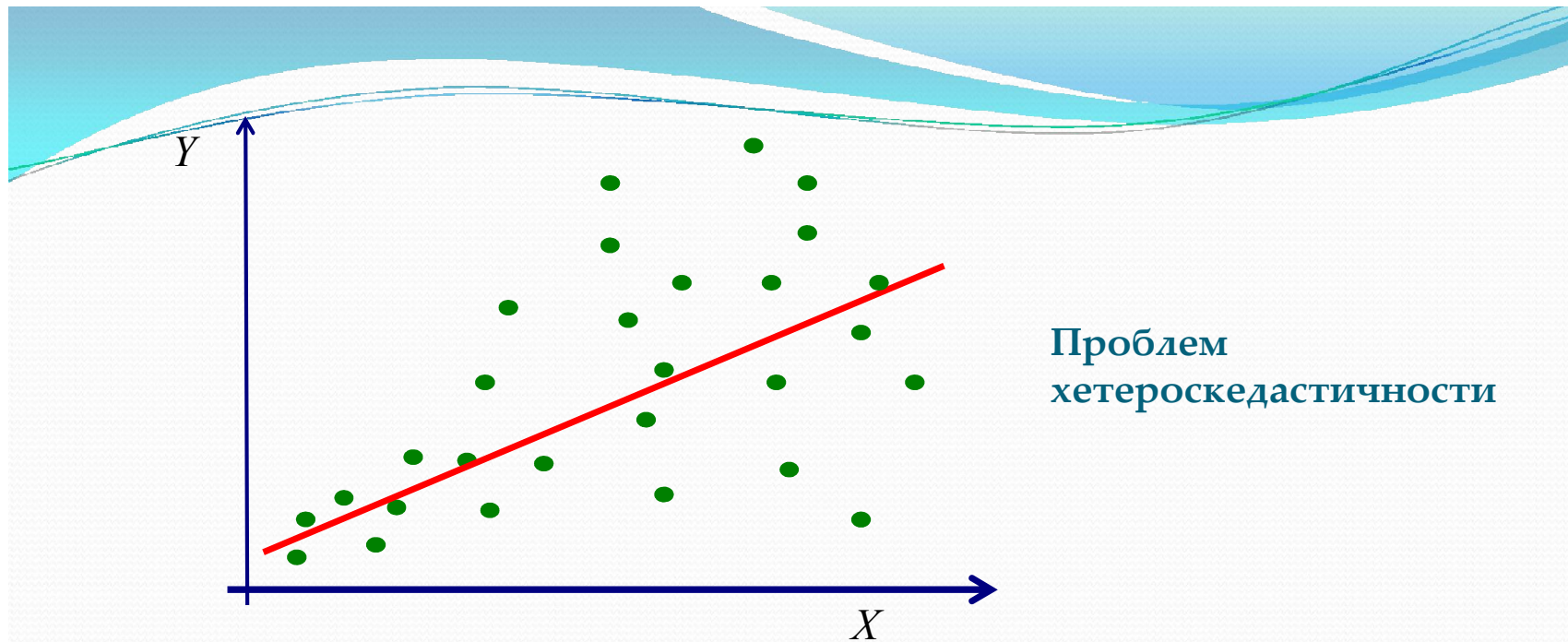
Пример 1. *Pooled* модел (результат из STATA софтвера)

. regress i c f

Source	SS	df	MS			
Model	4837650.66	2	2418825.33	Number of obs =	80	
Residual	1572709.03	77	20424.7926	F(2, 77) =	118.43	
Total	6410359.69	79	81143.7936	Prob > F =	0.0000	
				R-squared =	0.7547	
				Adj R-squared =	0.7483	
				Root MSE =	142.92	

i	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c	.3004544	.0493996	6.08	0.000	.2020871	.3988216
f	.1105221	.0137754	8.02	0.000	.0830916	.1379525
_cons	-62.83181	29.72504	-2.11	0.038	-122.0219	-3.641695

Претпоставка *pooled* модела – слободни члан не варира по јединицама (једнаки средњи нивои инвестиција све четири компаније).



Пример 1 – тестирање хетероскедастичности у *Pooled* моделу (оцењен методом ОНК)

```
. hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: fitted values of i
```

```
chi2(1)      =      22.52
```

```
Prob > chi2  =      0.0000
```

Пример 2. Модел са константним регр. параметрима - сектор мечног говедарства

```
. regress Nd Pr Sub Am Fvg region
```

Source	SS	df	MS	
Model	2.2309e+15	5	4.4619e+14	Number of obs = 350
Residual	5.7732e+14	344	1.6783e+12	F(5, 344) = 265.86
Total	2.8082e+15	349	8.0466e+12	Prob > F = 0.0000
				R-squared = 0.7944
				Adj R-squared = 0.7914
				Root MSE = 1.3e+06

Nd	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Pr	.7162428	.0277092	25.85	0.000	.6617419	.7707437
Sub	.6055271	.1641835	3.69	0.000	.2825973	.9284569
Am	-1.536081	.2018026	-7.61	0.000	-1.933003	-1.139159
Fvg	-3.2457	.1954146	-16.61	0.000	-3.630058	-2.861342
region	683929.4	226536.3	3.02	0.003	238358.8	1129500
_cons	-1376810	249236.8	-5.52	0.000	-1867030	-886590.6

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

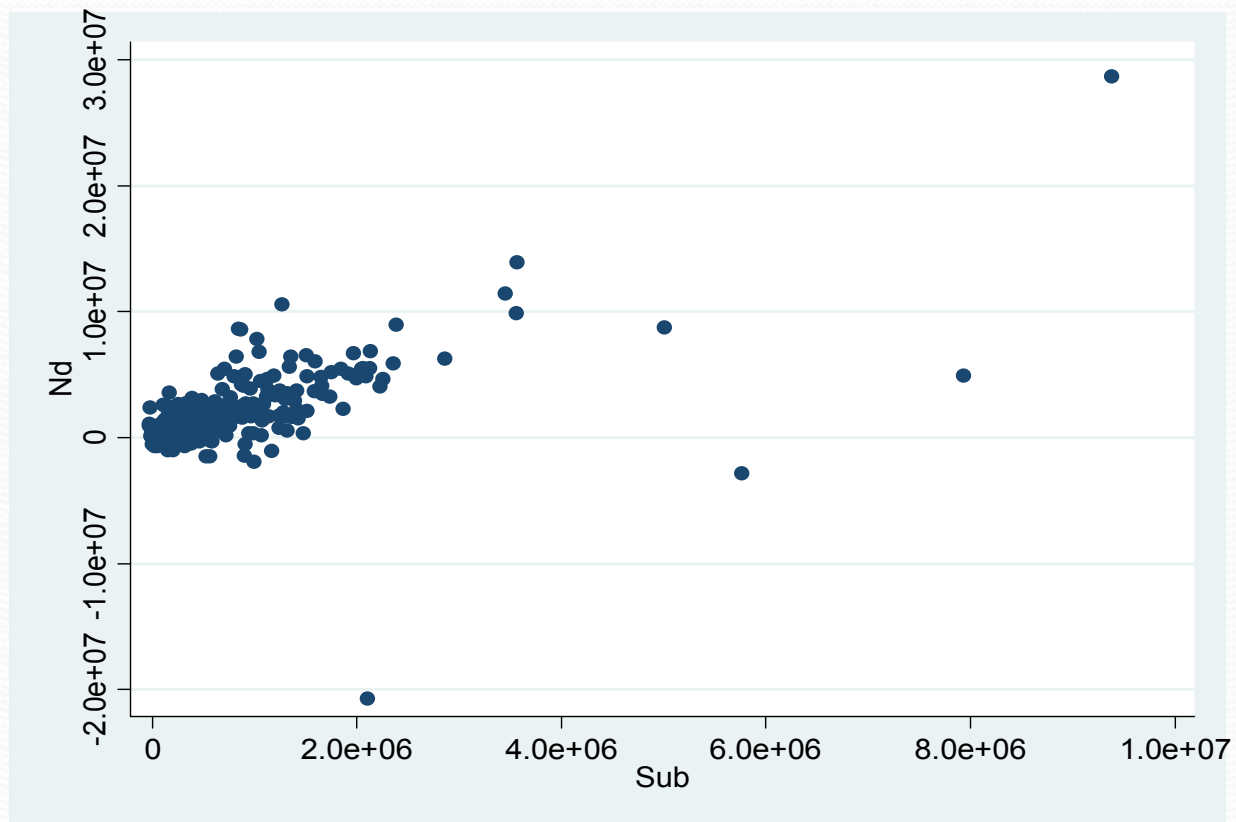
H0: Constant variance

Variables: fitted values of se420

chi2(1) = 16.08

Prob > chi2 = 0.0001

Пример 2: Субвенције и нето добитак пољопривредних газдинстава у сектору млекарског говедарства – присуство екстрема потенцијални узрок хетероскедастичности и неефикасних оцена



Модел са константним регресионим параметрима (*pooled model*)

- **Проблем оцењивања:** ако подаци панела показују велики варијабилитет, оцењивање *pooled* модела у пракси не решава проблем хетероскедастичности.
- **Решење:** уместо *pooled* модела оцењује се тзв. модел индивидуалних ефеката (хетерогеност се директно укључује у модел)

Пример 1. Модел са вештачким варијаблама

Модел без константе: број вештачких варијабли једнак је броју јединица посматрања

$$I_{it} = \beta_{11}D_1 + \beta_{12}D_2 + \beta_{13}D_3 + \beta_{14}D_4 + \beta_2F_{it} + \beta_2C_{it} + u_{it}$$

Модел са константом: број вештачких варијабли за један мањи од броја јединица посматрања

$$I_{it} = \beta_1 + \beta_{12}D_2 + \beta_{13}D_3 + \beta_{14}D_4 + \beta_2F_{it} + \beta_2C_{it} + u_{it}$$

D1=1 за GE,
=0 остало

D2=1 за GM,
=0 остало

D3=1 за US.Steel,
=0 остало

D4=1 за West
=0 остало

Пример 1. Модел са вештачким варијаблама оцењен методом ОНК -резултат из STATA софтвера-

```
. regress i c f gm us west
```

Source	SS	df	MS			
Model	5988160.47	5	1197632.09	Number of obs = 80		
Residual	422199.223	74	5705.39491	F(5, 74) = 209.91		
Total	6410359.69	79	81143.7936	Prob > F = 0.0000		
				R-squared = 0.9341		
				Adj R-squared = 0.9297		
				Root MSE = 75.534		

i	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c	.3450486	.0267082	12.92	0.000	.2918313	.3982658
f	.1084032	.0175653	6.17	0.000	.0734036	.1434027
gm	160.7042	46.62119	3.45	0.001	67.80942	253.5989
us	341.214	24.08021	14.17	0.000	293.2332	389.1949
west	186.8432	31.61873	5.91	0.000	123.8415	249.8449
_cons	-246.2304	35.93632	-6.85	0.000	-317.8351	-174.6257

Модел са вештачким варијаблама – модел фиксних индивидуалних ефеката (**FE модел** или **LSDV модел панела**).

Модел фиксних индивидуалних ефеката (ФЕ –енг. *fixed effects model*)

$$y_{it} = \beta_{1i} + \sum_{k=2}^K \beta_k x_{kit} + u_{it} \quad \text{где је} \quad \beta_{1i} = \beta_1 + \mu_i$$

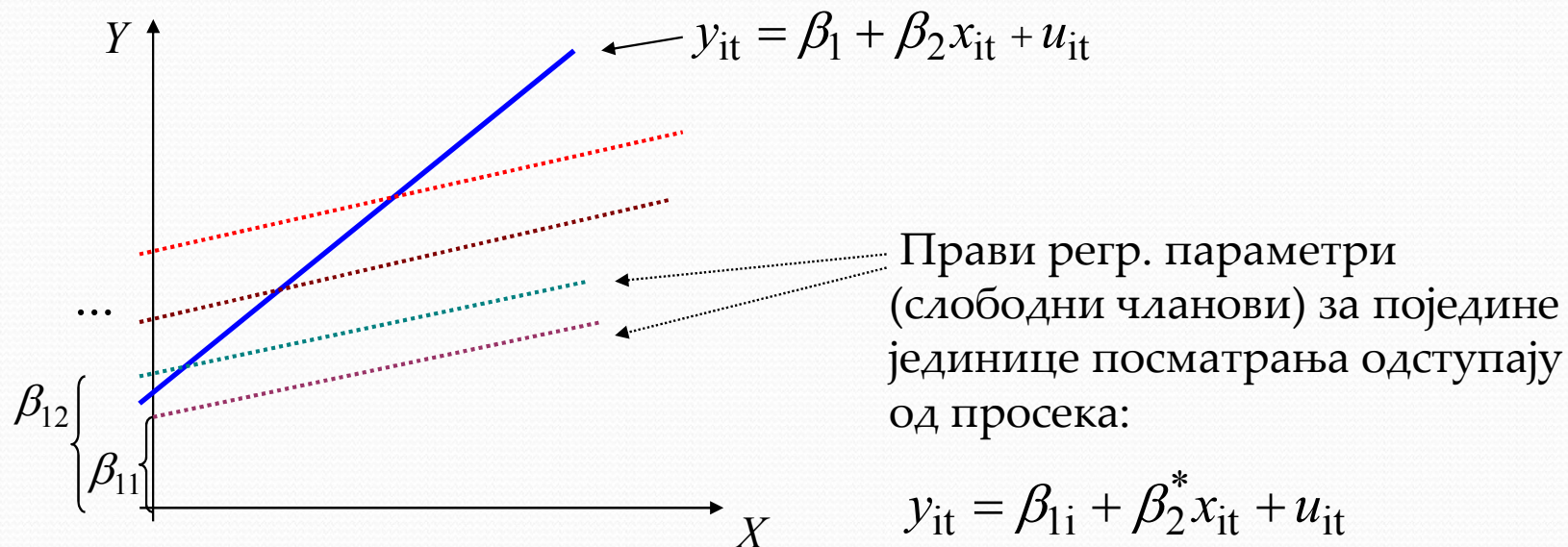
- β_1 – просек свих N јединица;
- μ_i – одступање β_{1i} i -те јединице посматрања од просека β_1
(**фиксни, непознати параметри**).
- Разлике (хетерогеност) између јединица посматрања преко варијабилних слободних чланова

μ_i - ефекти индивидуалних променљивих, које нису експлицитно укључене у модел, на варијације зависне променљиве по јединицама посматрања

Индивидуалне променљиве (енг. *time invariant variables*) имају различите вредности по јединицама, а исте током времена.

Модел фиксних индивидуалних ефеката

Занемаривање варијабилности регр. параметара (оцењивање уз претпоставку константних регресионих параметара) има за последицу неконзистентне оцене.



Оцењивање модела фиксних индивидуалних ефеката

- (1) Оцењивање ФЕ модела са укљученим вештачким променљивим примена метода ОНК (**LSDV метод**);
- најчешћа примена у случају мањег броја јединица посматрања.
- **Проблем оцењивања:** за велики број јединица N , губи се велики број степени слободе; инверзија матрице велике димензије $(N+K-1)$.

(2) Коваријациони метод (енг. *Covariance method*) - заснива се на тзв. центрираним вредностима (одступањима појединачних вредности од индивидуалног просека) :

$$(y_{it} - \bar{y}_i) = (\beta_1 - \bar{\beta}_1) + (\mu_i - \bar{\mu}_i) + \sum_{k=2}^K \beta_k (x_{kit} - \bar{x}_{ki}) + (u_{it} - \bar{u}_i)$$

$$(y_{it} - \bar{y}_i) = \sum_{k=2}^K \beta_k (x_{kit} - \bar{x}_{ki}) + (u_{it} - \bar{u}_i)$$

где су: \bar{y}_i , \bar{x}_{ki} и \bar{u}_i индивидуални просеци .

○ Примена LSDV метода и коваријационог метода на ФЕ модел даје **идентичне оцењене вредности коефицијената нагиба.**

LSDV метод

```
. regress i c f gm us west
```

Source	SS	df	MS			
Model	5988160.47	5	1197632.09	Number of obs =	80	
Residual	422199.223	74	5705.39491	F(5, 74) =	209.91	
Total	6410359.69	79	81143.7936	Prob > F =	0.0000	
				R-squared =	0.9341	
				Adj R-squared =	0.9297	
				Root MSE =	75.534	

i	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c	.3450486	.0267082	12.92	0.000	.2918313	.3982658
f	.1084032	.0175653	6.17	0.000	.0734036	.1434027
gm	160.7042	46.62119	3.45	0.001	67.80942	253.5989
us	341.214	24.08021	14.17	0.000	293.2332	389.1949
west	186.8432	31.61873	5.91	0.000	123.8415	249.8449
_cons	-246.2304	35.93632	-6.85	0.000	-317.8351	-174.6257

Коваријациони метод

```
. xtreg i c f, fe
```

Fixed-effects (within) regression
Group variable: **number**

R-sq: within = 0.8056
between = 0.7282
overall = 0.7534

corr(u_i, xb) = -0.1052

Number of obs = 80
Number of groups = 4
Obs per group: min = 20
avg = 20.0
max = 20

F(2,74) = 153.30
Prob > F = 0.0000

i	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c	.3450486	.0267082	12.92	0.000	.2918313	.3982658
f	.1084032	.0175653	6.17	0.000	.0734036	.1434027
_cons	-74.04004	37.65958	-1.97	0.053	-149.0784	.9982981



Тестирање индивидуалних и временских ефеката у фиксној спецификацији

- Избор спецификације модела панела зависи од резултата великог броја статистичких тестова.
- Тестирање индивидуалних и/или временских ефеката у моделу фиксних ефеката = тестирање варијабилности регресионих параметара по јединицама и / или кроз време.

Могућности примене панела у анализи економског положаја пољопривредних газдинстава у Србији

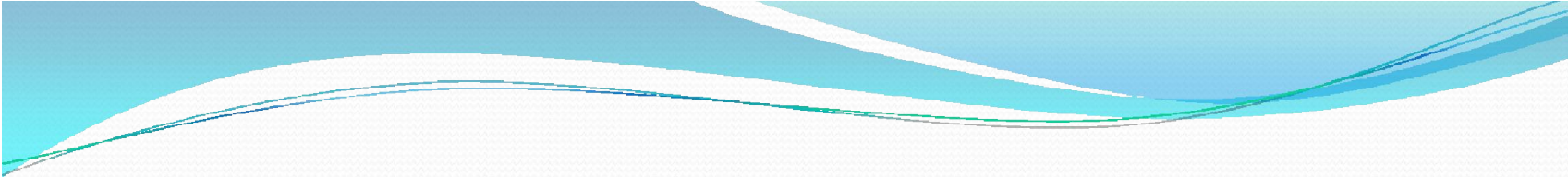
- Анализа ефеката детерминанти нето добитка пољопривредних газдинстава млечног говедарства:

$$ND_{it} = \beta_1 + \mu_i + \lambda_t + \beta_2 Pr_{it} + \beta_3 MP_{it} + \beta_4 Sub_{it} + \beta_5 Fvg_{it} + u_{it}$$

Корелациона матрица

	Nd	Pr	Sub	MP	Am	Fvg
Nd	1.0000					
Pr	0.7601	1.0000				
Sub	0.6043	0.8633	1.0000			
MP	0.4432	0.9139	0.8600	1.0000		
Am	0.1892	0.4725	0.4438	0.4816	1.0000	
Fvg	0.4417	0.8592	0.8478	0.9071	0.4185	1.0000

- Потенцијални проблем мултиколинеарности

- 
- Модел панела омогућава оцењивање:
 - различитих средњих вредности нето добитка по газдинствима (или групама газдинстава) и по годинама – индивидуални и временски ефекти, тј. варијабилни слободни чланови модела;
 - различитих ефеката изабраних фактора на нето добитак по регионима, годинама, величини нето додате вредности, приноса и тд. - варијабилни регр.коэффициенти нагиба...

ФЕ модел индивидуалних ефеката са робусним стандардним грешкама

Fixed-effects (within) regression
Group variable: gazd

Number of obs = 350
Number of groups = 175

R-sq: within = 0.9178
between = 0.6444
overall = 0.6090

Obs per group: min = 2
avg = 2.0
max = 2

corr(u_i, Xb) = -0.8915

F(5,170) = 379.48
Prob > F = 0.0000

Nd	Coef.	Std. Err.	t	P> t
Pr	1.007457	.0257096	39.19	0.000
Pr_r	-.3090213	.0664253	-4.65	0.000
Sub	.8211082	.2559846	3.21	0.002
Am	-.7322578	.4045575	-1.81	0.072
Fvg	-1.442941	.3462315	-4.17	0.000
_cons	-2180120	289081.3	-7.54	0.000
sigma_u	3759923.9			
sigma_e	720404.93			
rho	.96458905	(fraction of variance due to u_i)		

F test that all u_i=0: F(174, 170) = 5.57 Prob > F = 0.0000