اقتصاد سنجی داده های پانل

فصل هشت

غلامرضا کشاورز حداد

دانشگاه صنعتی شریف
دانشکده مدیریت و اقتصاد

9 أکتوبر 2021

فهرست مطالب

داده های های پانل

- در اين فصل واژه داده‌هاي پانل به ادغام مشاهدات صورت گرفته از داده‌هاي مقطعي مربوط به خانوارها، كشورها، بنگاه‌ها و ... گفته مي شود كه در طول دوره‌هاي زماني مختلف جمع‌آوري مي شود.
- اين داده‌ها با انجام بررسي‌هاي آماري و نمونه‌گيري‌هاي تصادفي يا سرشماري از خانوارها، بنگاهها و كشورهادر فاصله های زمانی منظم صورت مي‌گيرد، كه انجام مشاهدات نه تنها بر روي فرد بلكه همان افراد در طول زمان نيز شامل مي‌شود.
- از جمله مهمترين طرحهاي آماري كه در مركز آمار ايران بصورت پانل جمع‌آوري مي‌گردد.
  - نمونه‌گيري از وضعيت اقتصادي اجتماعي خانوارهاي شهري و روستايي در استانهاي مختلف به تفكيك مناطق شهري و روستايي براي ۳ سال در هر دهه است .
  - داده‌هاي مربوط به بعضي از شاخص‌هاي اقتصاد كلان مثل توليدناخالص داخلی و شاخص قيمتها در سطح استان‌ها نيز جمع‌آوري شده اند.
  - داده های بودجه خانوار ایران که به صورت پانل چرخشی از سال ۱۳۸۹ جمع آوری می گردد.

داده های های پانل

- هدف اين فصل فراهم ساختن يك مقدمه ساده از تكنيك داده‌هاي پانل به صورت معرفي مدل‌هاي OLS، اثرات ثابت و اثرات تصادفي است.
- براي رعايت سادگي و سهولت يادگيري، در اين فصل، مدل داده‌هاي پانل يك عاملی فردی، یک عاملی زمان و دو عاملی فرد و زمان به صورتهای اثرات ثابت و تصادفی با آزمونهای مشخص نمایی مربوطه معرفي مي شود.
- علاوه برآن مسئله درونزایی برخی از متغیرهای توضیحی در یک مدل پانل یک عاملی فردی نیز مورد بررسی قرار گرفته و چگونگی انجام آزمون اثرات تصادفی در مقابل اثرات فردی، در این مدلها نیز با ارایه یک مثال تجربی تشریح شده است.
- انجام تحقيقات ساده‌تر و آزمودن بعضي فرضيه‌هاي مطرح در علوم اجتماعي، اقتصاد و بطور كلي علوم رفتاري مي‌تواند پس از مطالعه اين فصل با بکارگیری نرم افزارهای مناسب به سادگي صورت گيرد.
- برآورد و تحليل مدل‌هاي پيشرفته تر اقتصاد سنجی مثل، مدلهاي با متغير وابسته محدود شده پانل، معادلات همزمان و داده‌هاي پانل پويا، در فصل ديگری معرفی می شود.

کاربرد تکنیکهای رگرسیونی پانل در دیگر رشته های علمی

- علاوه بر علوم اقتصادی، مدل رگرسيوني با داده‌هاي پانل موارد استفاده فراواني در ساير رشته‌هاي دانشگاه نيز دارد. بطور مثال:
  - در علوم سياسي، (بك و كاتز 1990)
  - در جامعه شناسي (انگلند و همكاران 1988)
  - در اقتصاد مالي (براون، كليدوني و ماتيس 1990)
  - در بازاريابي (اردم 1996)

- نگاهي به مجلات علمي مي‌تواند موارد كاربرد بيشتري را به اين فهرست اضافه نمايد. در ايران نيز اخيراً كاربردهاي قابل توجهي از اين تكنيك (اگرچه بعضاً با پاره‌اي از خطاها) در مجلات علمي ديده مي شود.
- اكنون اين پرسش جاي طرح دارد كه چه ضرورتي به استفاده از داده‌هاي پانل در تحقيقات علمي وجود دارد.

ضرورت استفاده از داده‌هاي پانل در تحقيقات علمي

- شااو (2003) . كلومارکن (1989) مزاياي زيادي براي استفاده از داده‌هاي پانل را فهرست مي‌كنند، که به چند مورد از آنها اشاره مي‌شود:
  - داده های پانل ناهمگني فردي را کنترل می کند. ايده اصلي مدلهاي داده پانل اين است که افراد، بنگاه‌ها، استانها و کشورها از نظر الگوي رفتاري همگن نيستند. مدلهاي داده‌هاي سري زماني به تنهايي و نيز داده‌هاي مقطعي به تنهايي، نمي‌توانندداين ناهمگني ها را مدلسازي نمايند. در نتيجه در صورت وجود ناهمگني تخمين‌زن‌هاي حاصل از آنها تورش دار ميشود.
  - داده‌هاي پانل، داده‌هاي با حجم زياد، با تغيير پذيري قابل توجه، هم خطي کم در ميان متغيرها، درجه آزادي بزرگتر و کارايي بيشتر را ايجاد مي‌كنند. مدلهاي سري زماني معمولاً داراي مشکل هم خطي هستند. بطور مثال در معادله رگرسيوني تقاضا براي سيگار هم خطي زيادي بين درآمد افراد و قيمت سيگار مصرفي آنها وجود دارد، اگر داده‌هاي سري زماني و مقطعي با همديگر بکار گرفته شود (داده‌هاي پانل) اين مشکل به اندازه چشمگيري بر طرف مي‌شود.

ضرورت استفاده از داده‌هاي پانل در تحقيقات علمي

- داده‌هاي پانل به نحو بهتري امكان مطالعه پويايي‌هاي تعديل را فراهم مي‌سازند. داده‌هاي مطالعات مقطعي نمي‌توانند به صورت پويا تصريح شوند، حال آنكه بكارگيري داده‌هاي سري زماني همراه با داده‌هاي مقطعي اين امكان را به سادگي فراهم مي‌سازند.
- دقت در داده ها. داده‌هاي خرد پانل جمع‌آوري شده بر روي افراد، بنگاه‌ها و خانوارها مي‌توانند با دقت بيشتري در مقايسه با متغير مشابه جمع‌آوري شده در سطح كلان، اندازه‌گيري و جمع‌آوري شوند، در نتيجه مدلهاي استوار يافته بر داده‌هاي پانل خرد فاقد تورش همفزوني هستند.
- حل مشكل كوتاه بودن دوره زماني داده ها. يكي از مشكلات داده‌هاي خرد مربوط به بنگاهها، خانوارها و افراد، اين است كه حجم مشاهدات انجام شده از نظر بعد زمان بسيار محدود است. ادغام داده هاي مقطعي و سري زماني مي‌تواند به مشكل حجم كم داده‌ها فائق آيد.

ساختار ماتریسی رگرسیون پانل

- یک مدل رگرسیونی پانل بطور همزمان می توانیم ناهمگنی (زمانی یا فردی) را در یک متغیر وابسته مطالعه کنیم.

$$y_{it} = \beta' \mathbf{x}_{it} + u_{it} \qquad t = 1, 2, ..., T \qquad i = 1, 2, ..., N$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ \vdots & \vdots \\ 1 & x_{1T} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{2T} \\ \vdots & \vdots \\ 1 & x_{N1} \\ 1 & x_{N2} \\ \vdots & \vdots \\ 1 & x_{NT} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ \vdots \\ u_{N1} \\ u_{N2} \\ \vdots \\ u_{NT} \end{bmatrix}; \quad \underset{(N.T)\times 1}{\mathbf{y}} = \alpha \underset{N.T\times 1}{\mathbf{l}} + \underset{(N.T)\times 1}{\mathbf{X}} \underset{1.1}{\beta} + \underset{(N.T)\times 1}{\mathbf{u}}$$

ساختار ماتریسی رگرسیون پانل

1. اگر u_{it} ها فرضهای رگرسیون کلاسیک خطی را برقرار سازند، آنگاه بکارگیری روش حداقل مربعات معمولی برآوردگرهای بدون تورش با کمترین واریانس را نتیجه می دهد، که به آن رگرسیون ادغام شده یا pooled می گوییم.
2. ممکن است این متغیر تصادفی فرضها کلاسیک را برقرار نسازند، آنگاه حالتهای زیر را خواهیم داشت:
   - مدل پانل یک عاملی فردی

$$u_{it} = \mu_i + \nu_{it}$$

   - مدل یک عاملی زمان

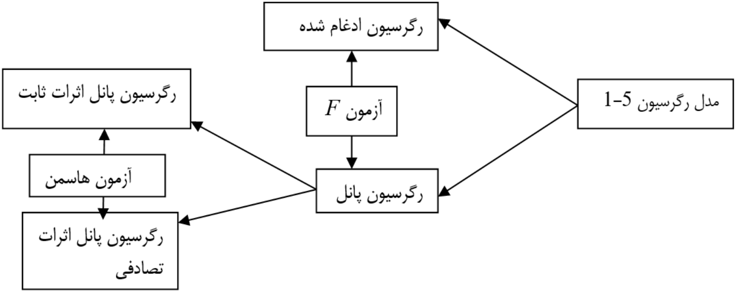
$$u_{it} = \lambda_t + \nu_{it}$$

   - دو عاملی فردی و زمان

$$u_{it} = \lambda_t + \mu_i + \nu_{it}$$

3. دو دسته فرض می توانیم برای μ_i و λ_t در نظر بگیریم:
   - μ_i و λ_t غیر تصادفی هستند
   - $\mu_i \sim i.i.d(0, \sigma_\mu^2)$ و $\lambda_t \sim i.i.d(0, \sigma_\lambda^2)$ تصادفی هستند

ساختار و فرایند مدلسازی پانل



شکل: اگر ناهمگنی بین افراد یا طول زمان اختلاف معنی داری از صفر داشته باشدمدل مورد مطالعه از ادغامی به پانل تبدیل می شود. در گام بعدی، اگر ناهمگنی از نوع اثر تصادفی نباشد، آن را به صورت اثر ات ثابت مدلسازی می کنیم.

رگرسیون ادغامی یا اثرات ثابت فرد؟

- در یک معادله رگرسیونی ساده پانل با یک متغیر توضیحی و عرض از مبدا داریم:

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \nu_{it} \quad t = 1, 2, ..., T \quad i = 1, 2, ..., N$$

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} x_{11} \\ \vdots \\ x_{1T} \\ x_{21} \\ \vdots \\ x_{2T} \\ \vdots \\ x_{N1} \\ \vdots \\ x_{NT} \end{bmatrix} \beta + \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_2 \\ \vdots \\ \mu_N \\ \vdots \\ \mu_N \end{bmatrix} + \begin{bmatrix} v_{11} \\ \vdots \\ v_{1T} \\ v_{21} \\ \vdots \\ v_{2T} \\ \vdots \\ v_{N1} \\ \vdots \\ v_{NT} \end{bmatrix}$$

رگرسیون ادغامی یا اثرات ثابت فرد؟

- که این دستگاه معادلات نشان مي‌دهند که μ_i ها پایا در زمان می‌باشند، مدل رگرسیونی تنها یک متغیر توضیحی x_{it} را داشته و β یک اسکالر است.
- در شکل برداری رابطه رگرسیونی پانل می‌تواند به صورت زیر نوشته شود:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{Z}_{\mu}\boldsymbol{\mu} + \boldsymbol{\nu}$$

- که درآن،

$$\mathbf{Z}_{\mu} = \mathbf{I}_N \bigotimes \mathbf{1}_T$$

Example

برای لحظه‌ای فرض را بر این قرار می‌دهیم که $N = 2$ و $T = 3$ باشد، آنگاه:

$$\mathbf{Z}_{\mu} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}'$$

رگرسیون ادغامی یا اثرات ثابت فرد؟

- با این تعریف، $\mathbf{Z}_\mu$ دربر گیرنده تعداد $N = 2$ متغیر مجازی است.
- بنابراین مدل پانل FE ممکن است با مشکل همخطی روبرو شود. چه می‌توان کرد؟
- **حذف عرض از مبدا یا حذف یکی از متغیرهای مجازی از مدل**

برآورد رگرسیون FE یا LSDV(within)

- با بکارگیری روش OLS برای رگرسیون H_a تخمین‌زنهای BLUE یا بهترین تخمین‌زنی خطی بدون تورش بدست می‌آید. ولي دو مشکل نه چندان چدی وجود دارد:
  - از دست دادن و کاهش درجه آزادی. زیرا $N+K$ پارامتر باید برآورد شود.
  - احتمال وجود همخطی و نیز یک ماتریس بزرگ $(\mathbf{X'X})$ که باید معکوس شود.
  - برای مثال اگر $N=50$ و $T=10$ و دو متغیر مستقل وجود داشته باشد، 500 مشاهده و 52 پارامتر وجود خواهد داشت.

- از نظر تکنیکی، راه حل بکار گیری رگرسیون *between* برای کاهش ابعاد ماتریس $(\mathbf{X'X})$است.

$$\bar{y}_{i.} = \alpha + \beta' \bar{\mathbf{x}}_{i.} + \mu_i + \bar{\nu}_{i.}$$

- که با میانگین‌گیری در طول زمان بدست می آید.
- ضرایب بدست آمده از این رگرسیون نشاندهنده اثر تغییر در x_{kit} از یک فرد به فرد دیگر در زمان خاص (نه در طول زمان برای یک فرد خاص) بر متغیر وابسته y_{it} را اندازه می گیرد.

برآورد رگرسیون با متغیرهای انحراف از میانگین بین مشاهدات $\bar{x}_{i.}$

- آنگاه رگرسیون $within$

$$y_{it} - \bar{y}_{i.} = \alpha + \beta'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) + \nu_{it} - \bar{\nu}_{i.}$$

با استفاده از روش حداقل مربعات معمولی برآورد میشود.

ساختار فنی داده ها در مدلهای پانل

	county	year	crmrte	prbarr
1	1	81	.0398849	.289696
2	1	82	.0383449	.338111
3	1	83	.0303048	.330449
4	1	84	.0347259	.362525
5	1	85	.036573	.325395
6	1	86	.0347524	.326062
7	1	87	.0356036	.29827
8	3	81	.0163921	.202899
9	3	82	.0190651	.162218
10	3	83	.0151492	.181586
11	3	84	.0136621	.194986
12	3	85	.0120346	.206897
13	3	86	.0129982	.156069

برآورد رگرسیون FE یا LSDV(within)

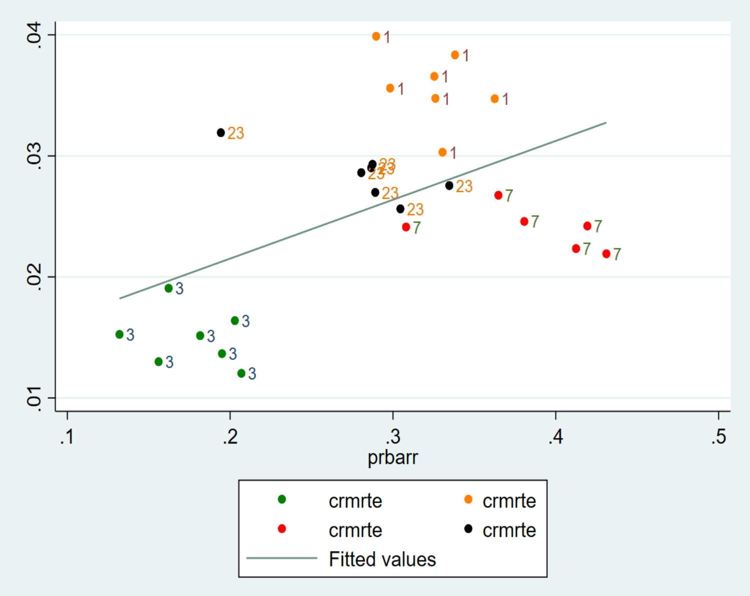
- از نظر تکنیکی، رگرسیون *pooled* به صورت زیر تعریف شده و با استفاده از *OLS* برآورد می شود.

$$y_{it} = \alpha + \beta' \mathbf{x}_{it} + u_{it}$$

VARIABLES	crmrte
prbarr	***0.0486
	(0.0167)
Constant	**0.0118
	(0.00502)
Observations	27
R-squared	0.253

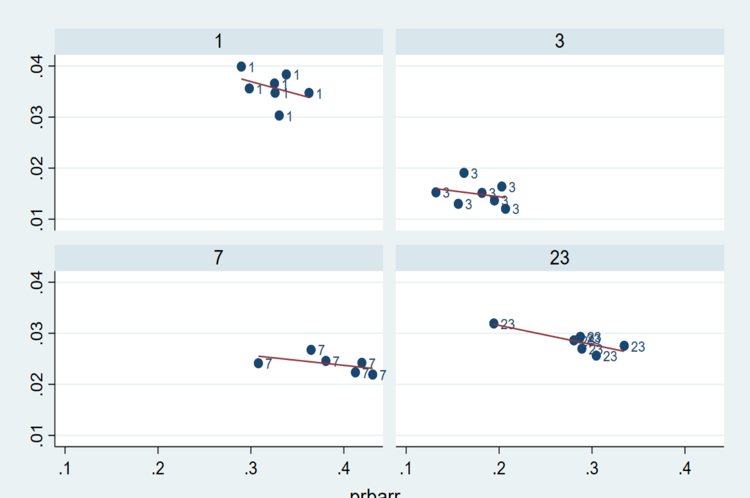
Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

بیان نموداری رگرسیون ادغامی



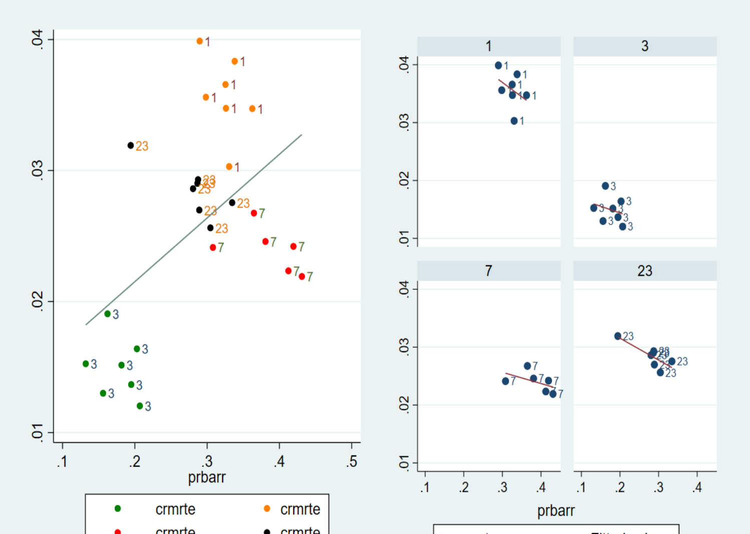
واقعیت چیست؟

- پراکنش داده های نرخ وقوع جرم در مقابل نرخ دستگیری، یک رابطه منفی را نشان می دهد. رگرسیون ادغامی ما را گمراه می کند!



دو نمودار در یک قاب

- برای سادگی مقایسه دو نمودار در یم قاب آورده می شود.



ریشه این تفاوت در مقدار تخمین زن ضرایب در چیست؟

- معمولا (نه همیشه) اثرات ثابت پایا در زمان مشاهده ناپذیر فردی μ_i با متغیرهای توضیحی مشاهده پذیر $\mathbf{x}_{it}$ همبستگی دارد، یعنی مشکل درونزایی بین متغیر توضیحی و جزء اخلال:

$$corr(x_{it}, \mu_i) \neq 0$$

- این همبستگی در پاره ای موارد آنچنان شدید است که سبب می شود تا رگرسیون ادغامی برآوردگرهایی ناسازگار، با مقدار عددی بزرگ ناسازگاری را نتیجه بدهد.
- به علامتهای منفی و مثبت این دو رگرسیون توجه کنید.

انتقال مبدا مختصات برای هریک از افراد به صفر

- آنگاه رگرسیون متغیرهای میانگین زدایی شده *within* عبارت است از:

```
bys county : egen mcrmrte=mean(crmrte)
bys county : egen mprbarr =mean( prbarr )
g dcrmrte = crmrte- mcrmrte
g dprbarr = prbarr- mprbarr
reg dcrmrte dprbarr
outreg2 using dmean, tex sideway replace
```

$$y_{it} - \bar{y_{i.}} = \alpha + \beta'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) + \nu_{it} - \bar{\nu_{i.}}$$

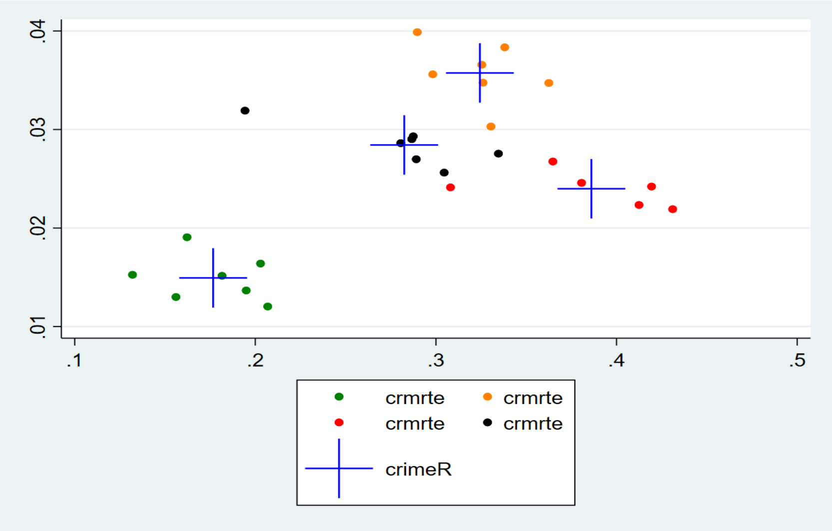
	(1)	(2)
VARIABLES	coef	se
dprbarr	**0.0305-	(0.0117)
Constant	-7.06e-10	(0.000387)
Observations	27	
R-squared	0.214	

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

بیان نموداری میانگین زدایی از داده ها، انتقال مبدا

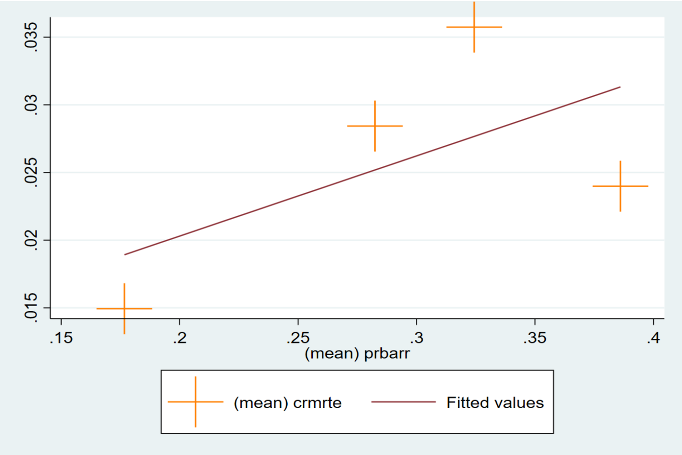
- رگرسیون میانگین $\bar{\mathbf{x}}_i.$ زدایی شده



رگرسیون بین افراد چیست؟

- رگرسیون *between*

$$\bar{y}_{i.} = \alpha + \beta' \bar{\mathbf{x}}_{i.} + \mu_i + \bar{\nu}_{i.}$$



رگرسیون بین افراد چیست؟

- رگرسیون *between*

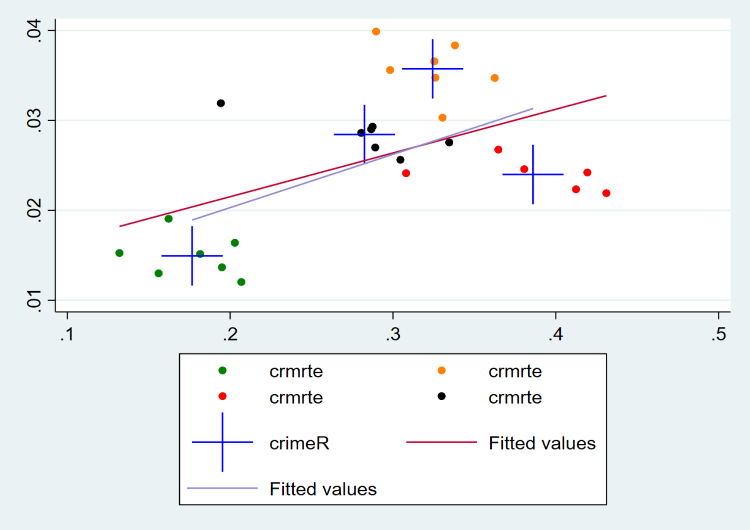
$$\bar{y}_{i.} = \alpha + \beta' \bar{\mathbf{x}}_{i.} + \mu_i + \bar{\nu}_{i.}$$

	(1)
VARIABLES	crmrte
prbarr	0.0592
	(0.0559)
Constant	0.00846
	(0.0169)
Observations	4
R-squared	0.360

Standard errors in parentheses

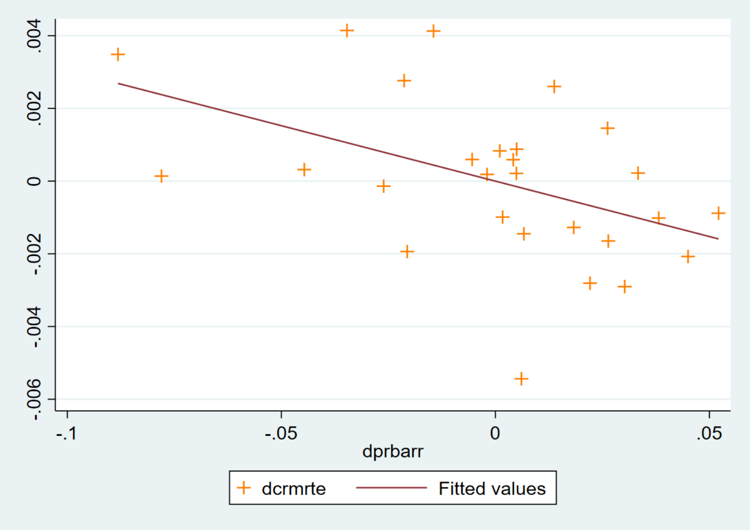
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

بیان نموداری رگرسیون ادغامی between و



کدام رگرسیون میتواند رابطه منفی دو متغیر را به تصویر بکشد؟

$$y_{it} - \bar{y}_{i.} = \alpha + \beta'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) + \nu_{it} - \bar{\nu}_{i.}$$



رگرسیون اثرات ثابت فردی

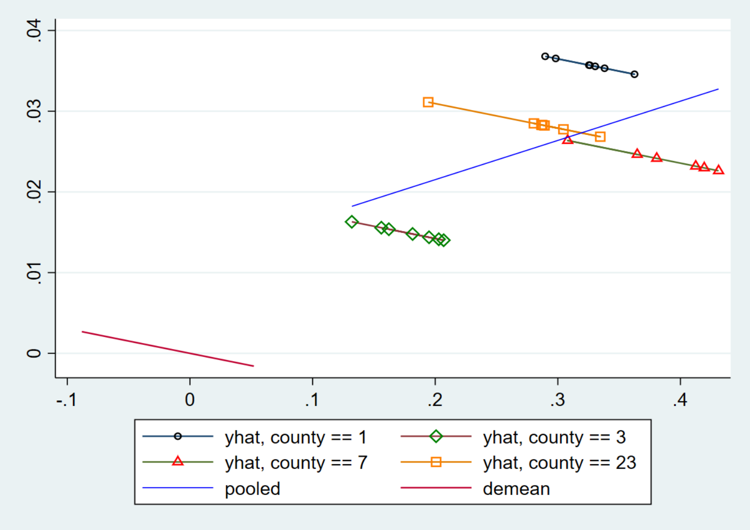
$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{Z}_{\mu}\boldsymbol{\mu} + \boldsymbol{\nu}$$

	(w.in, LSDV)	(2)	(pooled)	(4)
VARIABLES	crmrte-coef	se	crmrte-coef	se
crmrte				
prbarr	**0.0305_	(0.0124)	***0.0486	(0.0167)
3.county	***0.0253_	(0.00216)		
7.county	***0.00987_	(0.00142)		
23.county	***0.00859_	(0.00126)		
Constant	***0.0456	(0.00412)	**0.0118	(0.00502)
Observations	27		27	
R-squared	0.941		0.253	

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

کدام رگرسیون میتواند رابطه منفی دو متغیر را به تصویر بکشد؟



کد استاتا برای رگرسیونهای پانل

```
xtset county year, yearly
xtreg crmrte prbarr, fe
reg crmrte prbarr i.county
xtreg crmrte prbarr, be
```

VARIABLES	(FE) crmrte-coef	(2) se	(LSDV) crmrte-coef	(4) se
prbarr	**0.0305–	(0.0124)	**0.0305–	(0.0124)
3.county			***0.0253–	(0.00216)
7.county			***0.00987–	(0.00142)
23.county			***0.00859–	(0.00126)
Constant	***0.0347	(0.00362)	***0.0456	(0.00412)
Observations	27		27	
R-squared	0.214		0.941	
Number of county	4			

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

محدودیتهای رگرسیونهای اثرات ثابت

- مدل اثرات ثابت نمی توانند (در وضع موجود) عوامل مشاهده ناپذیر متغیر در طول زمان را کنترل کنند.
- از این نظر، هنوز درونزایی متغیرهای توضیحی با متغیرهای مشاهده ناپذیر ناپایا در زمان (مثل گرمایش زمین، ناکارآمدی نهادهای سیاسی،...) محتمل است.
- اندکی صبر کنید، برای این هم راه چاره داریم!

```
reg crmrte prbarr
outreg2 using pooledyear, sideway tex replace
xtset year county , yearly
xtreg crmrte prbarr, fe
outreg2 using pooledyear,  sideway tex append
reg crmrte prbarr i.year
outreg2 using pooledyear,  sideway tex append
xtreg crmrte prbarr, be
```

VARIABLES	crmrte pooled	crmrte FE	crmrte LSDV
prbarr	***0.0486	***0.0548	***0.0548
	(0.0167)	(0.0187)	(0.0187)
82.year			0.00222
			(0.00592)
83.year			0.00424_
			(0.00549)
84.year			0.00551_
			(0.00554)
87.year			0.000916_
			(0.00547)
Constant	**0.0118	*0.0100	*0.0122
	(0.00502)	(0.00559)	(0.00649)
Observations	27	27	27
R-squared	0.253	0.312	0.336
Number of year		7	

Standard errors in parentheses, year FE for 85 and 86 not reported

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

آزمون آماری رگرسیون ادغامی در مقابل پانل اثرات فردي

- می‌توان فرضیه،

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_{N-1} = 0$$

$$H_a :\sim H_0(\exists \mu_i \neq 0)$$

را با انجام یک آزمون F همانند آزمونهای نوع چو، یعني رگرسیون مقید(رگرسیون حداقل مربعات معمولی ساده) در مقابل رگرسیون غیر مقید (رگرسیون حداقل مربعات معمولی با متغیرهای مجازی افراد)، با استفاده از مجموع مجذورات پسماندها آزمون کرد.

$$\begin{array}{llll} \mathrm{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_{N-1} = 0, & & y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} & \rightarrow \mathbf{e}_r'\mathbf{e}_r \\ \mathrm{H}_a : \mu_i \neq \mu_j & (\exists\, \mu_i \neq 0), & y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \sum_{i=1}^{N-1} \mu_i D_{it} + v_{it} & \rightarrow \mathbf{e}_u'\mathbf{e}_u \end{array}$$

$$F = \frac{(\mathbf{e}_r'\mathbf{e}_r - \mathbf{e}_u'\mathbf{e}_u)/(N-1)}{\mathbf{e}_u'\mathbf{e}_u/(NT-K-N)} \sim F_{(N-1),\ (NT-K-N)}$$

جمع بندی رگرسیونهای اثرات ثابت

- اساسا بکارگیری اثرات ثابت در یک رگرسیون برای هر شهر، گروه، فرد و .... اثرات به طور مستقیم مشاهده ناپذیر این فاکتورها بر متغیر وابسته را برای هر فاکتور در طول زمان ثابت فرض کرده و برآورد می کند

- در نتیجه هدف به کارگیری متغیر دودویی اثرات ثابت، برای کنترل متغیرهای حذف شده پایا در زمان (استعداد، خوره کتاب بودن در دوره دانشجویی، داشتن معلم خوب در دوره تحصیل، تحصیل در دانشگاه خوب و ...) است.

- با این حال، هنوز ممکن است متغیرهای حذف شده دیگری وجود داشته باشند که ناپایا در زمان هستند، متغیرهای دودویی اثرات ثابت توانایی جانشینی برای این دسته از متغیرها را ندارند و لازم است از متغیرهای ابزاری استفاده کنیم. **این موضوع در بخشهای بعدی بررسی می شود.**

جمع بندی رگرسیونهای اثرات ثابت

- کنترل کردن برای این تفاوتهای بین فردی، تغییرات بین مقطعی مربوط به ناهمگنی های مشاهده ناپذیر مختص هر i را را حذف میکند.
- با استفاده از مدلهای اثرات ثابت ، اثرات متغیرهایی را که مقادیر آنها در طول زمان تغییر نمی کند ، **تخمین نمی زنیم**. در عوض ، **آنها را کنترل می کنیم و تأثیر آنها را از بین می بریم**.
- ***باقیمانده تغییرات (تغییرات برای هر فرد یا تغییرات طول زمان) برای شناسی یک رابطه علی بکار گرفته می شود، این دقیقا همان مفهومی است شما هر روز به آن فکر می کنید.***

چرا اثرات تصادفی؟ شروع یک داستان جدید

- هنگامی که متغیر توضیحی مورد نظر شما (مثلا یک متغیر سیاست) نسبت به اجزء خطا متعامد (دارای همبستگی صفر)است، باید از مدل RE استفاده کنید.
- بنابراین، در یک مدل اثرات تصادفی، فرض بر این است که متغیرهای مشاهده ناپذیر (پایا درزمان) با هیچ یک از متغیرهای مشاهده پذیر x_{kit} همبستگی ندارند (یا به بیان دقیق تر ، از نظر آماری مستقل هستند).
- اگر تردیدی وجود دارد (مثلا بنابه دلایل نظری) و فکر می کنید متغیرهای مستقل شما نسبت به جزء خطا متعامد نیست، از FE استفاده کنید.
- در مواردی که تجربه های تصادفی در چارچوب RTC تعیین می شود، می توان گفت که متغیرهای توضیحی کاملاً متعامد بوده و می توان ازبرآوردگر RE استفاده کرد.
- اگر چه، تقریباً تنها زمانی از برآوردگر RE استفاده می شود که داده های ما از یک تجربه تصادفی جمع آوری شده باشد، در بسیاری از موارد آزمونهای آماری هم به ما در پیدا کردن تصریح درست کمک می کنند. (در ادامه خواهیم دید)

رگرسیون پانل با اثرات تصادفی یک عاملی فردی

- اگر μ_i ها تصادفی و مستقل از متغیرهای توضیحی باشند، آنگاه مدل اثرات تصادفی یک مشخص‌نمایی مناسب برای مدلسازی رگرسیون پانل است.
- **رگرسیون پانل با اثرات تصادفی یک عاملی فردی را چگونه عملیاتی کنیم؟ آیا OLS را می توانیم بگار بگیریم؟ خیر، ولی می توانیم از GLS استفاده کنیم.**
- ماتریس واریانس کواریانس این مشخص‌نمایی با عناصر قطر اصلی $var(u_{it}) = \sigma_\mu^2 + \sigma_\nu^2$ برای تمام iها و tها واریانس همسان بوده ولی عناصر غیر قطر آن به صورت بلوکی است، که خودهمبستگی میان اجزاء اختلال یکسان را در طول زمان نشان می‌دهد:

$$
\begin{aligned}
\operatorname{Cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\nu^2 && i = j \quad s = t \\
&= \sigma_\mu^2 && i = j \quad s \neq t \\
&= 0 && i \neq j \quad s \neq t
\end{aligned}
$$

رگرسیون پانل با اثرات تصادفی یک عاملی فردی

$$E[u_{it} - \overbrace{E(u_{it})}^{0}][u_{js} - \overbrace{E(u_{js})}^{0}]$$

$$u_{it} = v_{it} + \mu_i$$

$$E(u_{it}) = E(v_{it} + \mu_i) = E(v_{it}) + E(\mu_i) = 0$$

برای $i = j, \quad s = t$

$$\begin{aligned} E[u_{it}u_{js}] &= E[u_{it}u_{it}] = E[v_{it} + \mu_i]^2 \\ &= E[v_{it}^2 + \mu_i^2 + 2\mu_i v_{it}] \\ &= E(v_{it}^2) + E(\mu_i^2) + 2E(v_{it}.\mu_i) = \sigma_v^2 + \sigma_\mu^2 \end{aligned}$$

به دلیل استقلال v_{it} و μ_i برای $s \neq t$ و $i = j$ داریم:

$$\begin{aligned} E(u_{it}u_{is}) &= E(\mu_i + v_{it})(\mu_i + v_{is}) \\ &= E(\mu_i^2 + \mu_i v_{is} + v_{it}\mu_i + v_{it}v_{is}) \\ &= E(\mu_i^2) = \sigma_\mu^2 \end{aligned}$$

رگرسیون پانل با اثرات تصادفی یک عاملی فردی

که در شکل ماتریسی، عبارتهای بالا به صورت ماتریس زیر قابل نمایش هستند.

$$
E(\mathbf{uu}')=\begin{bmatrix}
\sigma_\mu^2+\sigma_v^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 & \cdots & 0 & 0 & \cdots & 0 \\
\sigma_\mu^2 & \sigma_\mu^2+\sigma_v^2 & \cdots & \sigma_\mu^2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
\sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2+\sigma_v^2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & \sigma_\mu^2+\sigma_v^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\
0 & 0 & 0 & 0 & \cdots & \sigma_\mu^2 & \sigma_\mu^2+\sigma_v^2 & \cdots & \sigma_\mu^2 \\
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \cdots & \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2+\sigma_v^2
\end{bmatrix}
$$

$$\mathbf{u}=\mathbf{Z}_\mu.\boldsymbol{\mu}+\mathbf{V}$$

$$\boldsymbol{\Omega}=E\left[\mathbf{Z}_\mu.\boldsymbol{\mu}+\mathbf{V}\right]\left[\mathbf{Z}_\mu.\boldsymbol{\mu}+\mathbf{V}\right]'$$

رگرسیون پانل با اثرات تصادفی یک عاملی فردی

$$= E\left[\mathbf{Z}_{\mu}.\boldsymbol{\mu}\boldsymbol{\mu}'.\mathbf{Z}'_{\mu}\right] + \underbrace{E\left[\mathbf{Z}_{\mu}.\boldsymbol{\mu}\mathbf{V}'\right]}_{\mathbf{0}} + \underbrace{E\left[\mathbf{V}\boldsymbol{\mu}'.\mathbf{Z}'_{\mu}\right]}_{\mathbf{0}} + E\left[\mathbf{V}\mathbf{V}'\right]$$

$$= \sigma_{\mu}^{2}\left[\mathbf{I}_{N} \otimes \mathbf{J}_{TT}\right] + \sigma_{v}^{2}\left(\mathbf{I}_{N} \otimes \mathbf{I}_{T}\right)$$

$$= \begin{bmatrix} \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} & \cdots & 0 & 0 & \cdots & 0 \\ \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \cdots & \sigma_{\mu}^{2} \end{bmatrix} + \begin{bmatrix} \sigma_{v}^{2} & 0 & \cdots & 0 \\ 0 & \sigma_{v}^{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{v}^{2} \end{bmatrix}$$

رگرسیون پانل با اثرات تصادفی یک عاملی فردی

- مشاهده می شود که ماتریس واریانس کواریانس جز اخلال $\boldsymbol{u}$ قطری نبوده و این ابربردار دارای مشکل خودهمبستگی است.
- در یک جمع بندی می توان گفت که مدل رگرسیون پانل در صورت تصادفی بودن اثرات فردی(و نیز زمان)، پاره ای از فرضهای رگرسیون کلاسیک را برقرار نمی کند.
- به همین دلیل لازم است از تکنیک GLS استفاده شود.
- بردار تخمین زن GLS برای بردار پارامترهای β به صورت زیر بدست می آید.

$$\hat{\beta}_{RE} = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

- بردار تخمین زن LSDV یا within برای بردار پارامترهای β به صورت زیر بدست می آید.

$$\hat{\beta}_{FE} = [\mathbf{X}'\mathbf{Q}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y}$$

که در آن $\mathbf{Q} = \mathbf{I}_{N.T\times N.T} - \mathbf{I}_{N.N} \bigotimes \bar{\mathbf{J}}_{T.T}$ بوده و برای میانگین زدایی از متغیرهای رگرسیون پانل اثرات ثابت بکار بسته می شود. به خاطر بیاورید رگرسیونهای افراز شده را.

رگرسیون اثرات تصادفی در مقابل اثرات ثابت

- یک فرض خیلی مهم در مدل رگرسیونی داده های پانل این است که؛

$$E(u_{it}|x_{it}) = 0$$

این فرض در هنگامیکه جزء اختلال مدل رگرسیون پانل دربرگیرنده اثرات تصادفی پایا در زمان، غیر قابل مشاهده بوده و می‌تواند با x_{it}ها همبسته باشد، اهمیت بیشتری پیدا می‌کند.

- برای مثال، در یک معادله دریافتی این ها (مثلا هوش و استعداد فرد) می‌تواند با متغیرهایی نظیر سطوح تحصیلات (Schooling) که در سمت راست معادله رگرسیونی می‌آید همبسته باشد، در این حالت،

$$E(u_{it}|x_{it}) \neq 0$$

بوده و تخمین‌زنهای GLS، برای β تورش‌دار و ناسازگار خواهد بود.

- یعنی اگر متغیرهای توضیحی تحصیلات با متغیر حذف شده پایا در زمان استعداد همبسته باشد، تخمین زنهای GLS یا همان اثرات تصادفی ناسازگار خواهد بود.

رگرسیون اثرات تصادفی در مقابل اثرات ثابت

- با این حال، تخمین‌زن اثرات ثابت، $\boldsymbol{\beta}$ ها را ثابت و بجای متغیر توضیحی مشاهده ناپذیر مربوط به آنها، متغیرهای مجازي در نظر گرفته و در هنگام برآورد بردار $\boldsymbol{\beta}$ ، متغیرهای مدل پانل را μ زدایی می‌کند.
- در نتیجه این اقدام $\hat{\beta}_{FE}$ تخمین‌زن سازگار و بدون تورشی برای β می‌گردد.
- هاسمن (1978)، پیشنهاد مقایسه $\hat{\beta}_{RE}$ و $\hat{\beta}_{FE}$ را می دهد که هر دو آنها تحت فرضیه صفر

$$E(u_{it}|x_{it}) = 0$$

سازگار است، ولی در صورت عدم برقراری $E(u_{it}|x_{it}) = 0$ حد احتمال آنها متفاوت می‌شود.

- در واقع $\hat{\beta}_{FE}$ در هر دو صورت برقرار شدن $E(u_{it}|x_{it}) = 0$ یا برقرار نشدن آن سازگار است و در حالیکه $\hat{\beta}_{RE}$ تحت H_0 سازگار بطور مجانی کارآ و سازگار است، اما در صورت رد H_0 ناسازگار می‌باشد.

رگرسیون اثرات تصادفی در مقابل اثرات ثابت

- يعني اينكه، تعداد متغيرهاي توضيحي در مدل LSDV به اندازه تعداد افراد پانل منهاي يك ، بيشتر از مدل اثرات تصادفي است.
- افزايش تعداد متغيرهاي توضيحي باعث افزايش واريانس تخمين‌زن‌ها و در نتيجه كاهش كارآيي آنها مي‌شود.
- تحت فرضيه صفر هر دو تخمين‌زن سازگار مي‌باشند، اگرچه تحت فرضيه مقابل تخمين‌زن R.E ناسازگار ولي هنوز تخمين‌زن‌هاي LSDV سازگار هستند.
- بنابراين ما در درجه اول علاقه‌مند به داشتن تخمين‌زن‌هاي سازگار و كارآ (اثرات تصادفي) هستيم، در غير اينصورت (رد فرضيه صفر) تخمين‌زن با كارآيي كمتر (اثرات ثابت) به تخمين‌زن‌هاي ناسازگار (اثرات تصادفي) ترجيح داده مي‌شود.

رگرسیون اثرات تصادفی در مقابل اثرات ثابت

- آماره هاسمن برای آزمودن درستی رگرسیون اثارت تصادفی در مقابل اثرات ثابت به صورت زیر تعریف می شود.

$$Wald = \hat{\mathbf{q}}'\left[\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{F.E.}}\right) - \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right)\right]^{-1}\hat{\mathbf{q}} \sim \chi^2_K$$

که تحت فرضیه H_0 به این تابع نمونه‌ای دارای توزیع مجانبی χ^2_K است که در آن K بعد بردار ضرایب $\boldsymbol{\beta}$ ست. برای عملیاتی ساختن این آزمون، تخمین‌زن سازگار شدنی $\hat{\boldsymbol{\Omega}}$ جایگزین $\boldsymbol{\Omega}$ می‌شود.

- که در آن $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$ است.

یک نکته تکنیکی درباره ماتریس واریانس ـ کواریانس $\hat{\mathbf{q}}$

- استخراج واریانس ـ کواریانس $\hat{\mathbf{q}}$

$$W = \hat{\mathbf{q}}'\left[\mathrm{V.C}\left(\hat{\mathbf{q}}\right)\right]^{-1}\hat{\mathbf{q}} \sim \chi^2_{\mathrm{Rank(V.C(\hat{\mathbf{q}}))}}$$

$$\mathrm{V.C}\left(\hat{\mathbf{q}}\right) = ?$$

$$\mathrm{V.C}\left(\hat{\mathbf{q}}\right) = \mathrm{var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{F.E.}}\right) - \mathrm{var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{GLS.}}\right) \quad !!!$$

می‌دانیم تحت فرضیه صفر $plim\ \hat{q} = 0$

$$\mathrm{H}_0 : plim\ \hat{\mathbf{q}} = plim(\tilde{\boldsymbol{\beta}}_{\mathrm{FE}} - \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = 0$$

چون در صورت نبود درونزایی متغیرهای توضیحی، هر دو تخمین‌زن اثرات ثابت و تصادفی در احتمال به $\boldsymbol{\beta}$ می‌گرایند (یا سازگار هستند).

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} - \boldsymbol{\beta} = \left[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{u}$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{F.E.}} - \boldsymbol{\beta} = \left[\mathbf{X}'\mathbf{Q}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Q}\mathbf{u}$$

$$\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_{\mathrm{F.E.}} - \hat{\boldsymbol{\beta}}_{\mathrm{GLS}} \rightarrow \hat{\boldsymbol{\beta}}_{\mathrm{F.E.}} = \hat{\mathbf{q}} + \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$$

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{F.E.}}\right) = \mathrm{var}\left(\hat{\mathbf{q}}\right) + \mathrm{var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right) + 2\mathrm{cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \hat{\mathbf{q}}\right)$$

یک نکته تکنیکی درباره ماتریس واریانس- کوواریانس$\hat{\mathbf{q}}$

- استخراج واریانس- کوواریانس $\hat{\mathbf{q}}$. ادامه

$$\hat{\boldsymbol{\beta}}_{\text{F.E.}} - \boldsymbol{\beta} = \left[\mathbf{X}'\mathbf{Q}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Q}\mathbf{u}$$

$$\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_{\text{F.E.}} - \hat{\boldsymbol{\beta}}_{\text{GLS}} \rightarrow \hat{\boldsymbol{\beta}}_{\text{F.E.}} = \hat{\mathbf{q}} + \hat{\boldsymbol{\beta}}_{\text{GLS}}$$

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{F.E.}}\right) = \operatorname{var}\left(\hat{\mathbf{q}}\right) + \operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}\right) + 2\operatorname{cov}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}, \hat{\mathbf{q}}\right)$$

$$\begin{aligned}
\operatorname{cov}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}, \hat{\mathbf{q}}\right) &= E\left(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right)\left(\hat{\mathbf{q}} - E\left(\hat{\mathbf{q}}\right)\right)' = E\left(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right)\left(\hat{\mathbf{q}}\right)' \\
&= E\left[\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right]\left[\left(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}\right) - \left(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right)\right]' \\
&= E\left\{-\left[\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right]\left[\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right]' + \left[\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}\right]\left[\hat{\boldsymbol{\beta}}_{\text{F.E.}} - \boldsymbol{\beta}\right]'\right\} \\
&= -\operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}\right) + E\left[\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\underbrace{\mathbf{u}\mathbf{u}'}_{\boldsymbol{\Omega}}\mathbf{Q}\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{X}\right)^{-1}\right] \\
&= -\operatorname{var}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) + [(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\underbrace{\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}}_{\mathbf{I}}\mathbf{Q}\mathbf{X}(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}]
\end{aligned}$$

یک نکته تکنیکی درباره ماتریس واریانس ـ کوواریانس $\hat{\mathbf{q}}$

- استخراج واریانس ـ کوواریانس $\hat{\mathbf{q}}$. ادامه

$$= -\mathrm{var}(\hat{\boldsymbol{\beta}}_{GLS}) + [(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{Q}\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{X}\right)^{-1}}_{\mathbf{I}}]$$

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}_{GLS}, \hat{\mathbf{q}}\right) = -\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) + \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1} = \mathbf{0}$$

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = [(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\underbrace{\mathbf{u}\mathbf{u}'}_{\boldsymbol{\Omega}}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}]$$

$$= [(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\underbrace{\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}}_{\mathbf{I}}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}]$$

$$= \left[\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\underbrace{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}}_{\mathrm{I}}\right]$$

$$= \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

یک نکته تکنیکی درباره ماتریس واریانس ـ کواریانس $\hat{\mathbf{q}}$

- استخراج واریانس ـ کواریانس ـ کواریانس $\hat{\mathbf{q}}$. ادامه

$$\operatorname{cov}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}, \hat{\mathbf{q}}\right) = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1} - \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1} = \mathbf{0}$$

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{F.E.}}\right) = \operatorname{var}\left(\hat{\mathbf{q}}\right) + \operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}\right)$$

$$\Rightarrow \operatorname{var}\left(\hat{\mathbf{q}}\right) = \operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{F.E.}}\right) - \operatorname{var}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}\right)$$

از اینرو تابع نمونه‌ای آزمون هاسمن، یک تابع نمونه‌ای والد است که به صورت زیر تعریف می‌شود:

$$Wald = \hat{\mathbf{q}}'\left[\operatorname{Var}\left(\hat{\boldsymbol{\beta}}_{\text{F.E.}}\right) - \operatorname{Var}\left(\hat{\boldsymbol{\beta}}_{\text{GLS}}\right)\right]^{-1}\hat{\mathbf{q}} \sim \chi^2_K$$

که تحت فرضیه H_0 به این تابع نمونه‌ای دارای توزیع مجانبی χ^2_K است

که در آن K بعد بردار ضرایب $\boldsymbol{\beta}$ است. برای عملیاتی ساختن این آزمون،

تخمین‌زن سازگار شدنی $\hat{\boldsymbol{\Omega}}$ جایگزین $\boldsymbol{\Omega}$ می‌شود.

آزمون هاسمن(فرضیه اثرات تصادفی در مقابل اثرات ثابت)

```
xtreg wage schooling experience , fe
est store fereg
xtreg wage schooling experience , re
est store rereg
hausman fereg rereg
```

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
  schooling_ |    .0713003    -.1942186        .2655188        .0174657
 experience_ |    .0068309     .016212        -.0093812         .001654
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      244.09
                Prob>chi2 =      0.0000
```

- **قضاوت:** اگر مقدار تابع نمونه‌اي محاسبه شده از کوانتیل چي‌دو جدول آماري بزرگتر باشد، فرضیه صفر مبني بر درست بودن مشخص‌نمايي اثرات تصادفي رد مي‌شود. بنابراین لازم است از مشخص نمايي اثرات ثابت استفاده شود.

یک نکته تکنیکی درباره ماتریس واریانس- کواریانس$\hat{\mathbf{q}}$

- ایا ماتریس واریانس- کواریانس$\hat{\mathbf{q}}$ همیشه معکوس پذیر است؟ **خیر**
- راه حل چیست؟ استفاده از تعمیم معکوس سازی ماتریس به روش Moore–Penrose آستفاده می کنیم.

$$Wald = \hat{\mathbf{q}}'\left[\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{F.E.}}\right) - \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right)\right]^{-}\hat{\mathbf{q}} \sim \chi^2_K$$

```
xtreg crmrte prbarr , fe
est store fereg
xtreg crmrte prbarr , re
est store rereg
hausman fereg rereg , sigmamore
```

آزمون هاسمن(فرضیه اثرات تصادفی در مقابل اثرات ثابت)

```
. hausman fereg rereg

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fereg        rereg        Difference          S.E.
-------------+----------------------------------------------------------------
      prbarr |   -.0304908    -.0262502      -.0042406               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =   -13.93    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

. hausman fereg rereg, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fereg        rereg        Difference          S.E.
-------------+----------------------------------------------------------------
      prbarr |   -.0304908    -.0262502      -.0042406        .0027821
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.32
                Prob>chi2 =      0.1274
```

تفسیر نتایج آزمون فرضیه اثرات تصادفی در مقابل اثرات ثابت

- خروجی نخست نشان می دهد که مقدار عددی آماره والد محاسبه شده کوچکتر از صفر بود و داده های مورد استفاده در مدل فرضهای مجانبی این آماره آزمون را برقرار نمی سازنند.

- خروجی دوم با بکارگیری گزینه sigmamore در دستور آزمون هاسمن بدست آمده است.
  - مقدار آماره محاسبه شده برابر با ۲,۳۲ ، با مقدار احتمال محاسبه شده ۰,۱۲۷۴ برای آزمون فرضیه صفر (نه مقداراحتمال ناحیه بحرانی آزمون) بزرگتراز مقدار مرسوم ۵درصد است.
  - درنتیجه نمی توانیم فرضیه صفر درستی مشخص نمایی اثرات تصادفی را رد کنیم.

مدلهای اثرات تصادفی، اثرات ثابت و ادغام شده در یک قاب

VARIABLES	F.E crmrte	R.E crmrte	Pooled crmrte
prbarr	**0.0305-	**0.0263-	***0.0486
	(0.0124)	(0.0125)	(0.0167)
Constant	***0.0347	***0.0334	**0.0118
	(0.00362)	(0.00571)	(0.00502)
Observations	27	27	27
R-squared	0.214		0.253
Number of county	4	4	

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

- در نمونه آماری ما، برخلاف مدل ادغامی، دو مدل اثرات تصادفی و ثابت ضرایب برآورد شده همسویی دارند، و فاصله اعتماد هردو آنها تخمین نقطه ای دیگری را دربرمی گیرد، یعنی اختلاف معنی درای از هم ندارد.

مدل رگرسيوني پانل با ضرايب متغير

- در بخش‌هاي پيش با موردي از مدل‌هاي رگرسيوني پانل روبر بوديم که تنها عرض از مبدأ مدل رگرسيون در ميان افراد (يا زمان يا هر دو اينها) متفاوت بودند.

- در بخش حاضر بر مدلهایی متمرکز می شویم که در آنها شیب متغیرهای توضيحي در ميان افراد متفاوت است.
  - بطور مثال با توجه به ساختارهاي تکنولوژيك و سطح توسعه يافتگي استان‌ها، واکنش رشد اقتصادی در استان به متغیر مخارج عمرانی دولت میتواند متفاوت باشد، يا اينکه واکنش تقاضاي برق استان‌هاي گرمسير به تغييرات قيمت برق و نيز واکنش مصرف نفت خام کشورهاي توسعه يافته و جهان سوم به خاطر تفاوت در ساختار تکنولوژيك، نسبت به تغييرات قيمت نفت خام متفاوت است.

- در نتيجه مي‌توان مدل‌هاي رگرسيوني پانلي را طراحي کرد که در آن نه تنها عرض از مبدأ در ميان افراد متفاوت است، بلکه ضرايب تمامی يا بخشي از متغيرهاي توضيحی نیز در ميان افراد مي‌تواند متفاوت باشد.

مدل رگرسيوني پانل با ضرايب متغير

- در اين بخش، تنها به تفاوت در ميان افراد تمركز مي‌كنيم و فرض را براين قرار می دهيم كه اين ضرايب در طول زمان پايا و غير تصادفی هستند.

- اگرچه ضريب متغيرهای توضيحی در طول زمان نيز می تواند متفاوت باشد، در اين بخش تنها به معرفی رگرسيونهای پانل با ضرايب متفاوت(بين افراد) تمركز می كنيم.

- مدل رگرسيوني پانل با ضرايب متفاوت در بين افراد را در نظر بگيريد: (10)

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\beta}_i + u_{it}; \quad i = 1,\ldots,N, \quad t = 1,\ldots,T \quad \text{يا} \quad \mathbf{y}_i = \mathbf{z}_i\boldsymbol{\delta}_i + \mathbf{u}_i; \quad i = 1,\ldots,N$$

كه در آن $\mathbf{y}'_i = [y_{i1} \quad ,\ldots, \quad y_{iT}]$ و $\mathbf{z}_i = [\mathbf{i}_T \vdots \mathbf{x}_i]$ و $\mathbf{x}_i$ يک ماتريس $T \times K$،
بردار پارامترهای $\boldsymbol{\delta}'_i$ دارای ابعاد $1 \times (K+1)$ و $\mathbf{u}_i$ دارای بعد $T \times 1$ است.

- نكته حائز اهميت اين است كه ضرايب δ_i در معادله رگرسيوني بالا، برای افراد (استان، كشور، منطقه و ...) متفاوت است.

مدل رگرسيوني پانل با ضرايب متغير

- هدف ما برآورد δ_i در معادله رگرسيون بالا است، كه در آن ضرايب تمام متغيرهاي توضيحي در بين افراد پانل تغيير پذير باشد،

- حال موضوع اصلی اين است كه چگونه مي‌توان تخمين زن‌هاي سازگاري را براي بردارهاي $\delta_i; \quad i = 1, \ldots, N$ در معادله رگرسيون (۱۰)بدست آورد.

- ابتدا معادله رگرسيون (۱۰) را به صورت ابر بردار و ابرماتريس می نويسيم.

$$\mathbf{y} = \mathbf{Z}^* \boldsymbol{\delta} + \boldsymbol{u}$$

- به سادگی مي‌توان نتيجه گرفت كه تخمين زنهاي حداقل مربعات معمولی به صورت

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}^{*\prime} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\prime} \mathbf{y}$$

بدست مي‌آيد، كه در آن ضريب متغيرهاي توضيحي و عرض از مبدأ مدل رگرسيوني براي افراد يكسان نيستند.

مدل رگرسيوني پانل با ضرايب متغير

- در اين مدل نيز چگونگي انجام استنتاج آماري درباره معني دار بودن ضرايب، همانند مدل رگرسيون پانل با عرض از مبدأهای متفاوت است.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_N \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix} = \mathbf{Z}^* \boldsymbol{\delta}^* + \mathbf{u}$$

```
separate prbarr , by(county)
levelsof county, local(countys)
foreach i of local countys {
replace prbarr`i'=0 if prbarr`i'==.
}
reg crmrte prbarr1-prbarr23 i.county
```

برآورد مدل اثرات تصادفي و متغير توضيحي درونزا

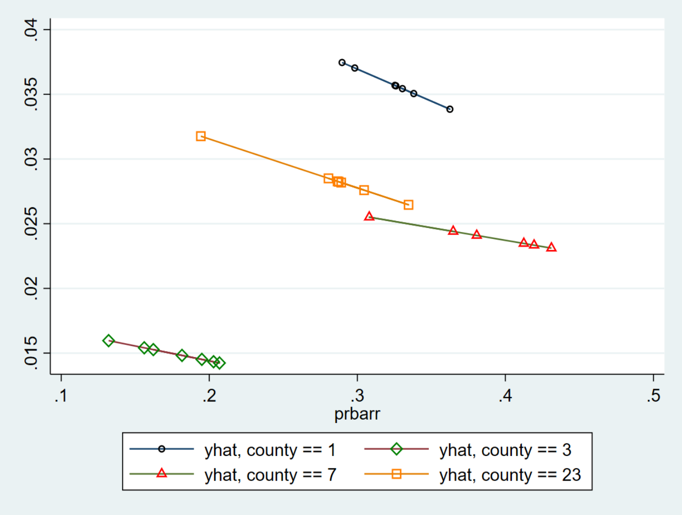
```
. reg crmrte prbarr1-prbarr23 i.county

      Source |       SS           df       MS      Number of obs   =        27
-------------+----------------------------------   F(7, 19)        =     45.02
       Model |  .001617019         7  .000231003   Prob > F        =    0.0000
    Residual |  .000097498        19  5.1315e-06   R-squared       =    0.9431
-------------+----------------------------------   Adj R-squared   =    0.9222
       Total |  .001714517        26  .000065943   Root MSE        =    .00227

------------------------------------------------------------------------------
      crmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     prbarr1 |  -.0495707   .0379184    -1.31   0.207    -.1289348    .0297935
     prbarr3 |   -.023043   .0334457    -0.69   0.499    -.0930458    .0469597
     prbarr7 |  -.0194486   .0222084    -0.88   0.392    -.0659312    .0270341
    prbarr23 |  -.0378029   .0215289    -1.76   0.095    -.0828634    .0072575
             |
      county |
          3  |  -.0328127   .0136985    -2.40   0.027     -.061484   -.0041413
          7  |  -.0203221   .0150456    -1.35   0.193    -.0518129    .0111688
         23  |  -.0127104   .0137737    -0.92   0.368     -.041539    .0161183
             |
       _cons |     .05182   .0123289     4.20   0.000     .0260153    .0776247
------------------------------------------------------------------------------
```

مدل رگرسيوني پانل اثرات ثابت با ضرايب متغير

- بیان نموداری خروجی رگرسیون پانل اثرات ثابت فردی با ضرایب متغیر



برآورد مدل اثرات ثابت و متغير توضيحي درونزا

- در بخش نخست تنها به مشكل درونزايي متغيرهاي توضيحي در يك مدل داده‌هاي پانل با اثرات ثابت یک عاملی فردی، پرداخته مي‌شود.

- متغیر تجربه کاری در معادله دستمزد را به خاطر آورید.

- هر اندازه دستمزد دریافتی فرد در یک شغل بیشتر باشد انگیزه بیشتری برای ادامه همکاری در آن شغل خواهد داشت، این یعنی درونزایی متغیر تجربه کاری در آن شغل.

- تعداد متغیرهای توضیحی درونزا میتواند بیش از یک مورد باشد، این بستگی به دیتا و متغیرهای توضیحی مدل تصریح شده دارد.

- دراینجا هم می توانیم از تکنیک برآوردگرهای متغیرهای ابزاری استفاده کنیم

تقسیم بندی مدلهای رگرسیونی با درونزايي متغيرهاي توضيحي

- آیا تورش ناشی از متغیرهای حذف شده از مدل تنها به دلیل حذف متغیرهای پایا در زمان بوده و $E(\mu_i|x_{it}) \neq 0$ است؟
  - **روش حل مشکل:** مدل متعارف اثرات ثابت

- آیا تورش ناشی از متغیرهای حذف شده از مدل به دلیل حذف متغیرهای پایا در زمان و نیز ناپایا در زمان بوده و $E(\mu_i|x_{it}) \neq 0$ است؟
  - **روش حل مشکل:** مدل متعارف اثرات ثابت بعلاوه تکنیک رگرسیون متغیر ابزاری

- آیا متغیرهای حذف شده از مدل تنها از نوع متغیرهای پایا در زمان بوده ولی $E(\mu_i|x_{it}) = 0$ است؟
  - **روش حل مشکل:** مدل متعارف اثرات تصادفی

- آیا تورش ناشی از متغیرهای حذف شده از مدل به دلیل حذف متغیرهای ناپایا در زمان بوده ولی $E(\mu_i|x_{it}) = 0$ است؟
  - **روش حل مشکل:** مدل متعارف اثرات تصادفی به علاوه متغیرهای ابزاری

برآورد مدل اثرات ثابت و متغیر توضيحي درونزا

- در بخش نخست تنها به مشكل درونزايي متغيرهاي توضيحي در يك مدل داده‌هاي پانل با اثرات ثابت یک عاملی فردی، پرداخته مي‌شود.

- مدل تک معادله‌ای $y_{it} = \mathbf{x}'_{1it}\boldsymbol{\beta} + \mathbf{y}'_{2it}\boldsymbol{\gamma} + u_{it}$ را در نظر بگیرید که در آن $\mathbf{y}_{2it}$ بردار متغیرهای توضیحی درونزا و $\mathbf{x}_{1it}$ بردار متغیرهای توضیحی برونزا بوده و در شکل ماتریسی به صورت زیر نوشته می‌شود:

$$\mathbf{y} = [\mathbf{X}_1 \quad \mathbf{Y}_2][\boldsymbol{\beta}' \quad \boldsymbol{\gamma}']' + \mathbf{u} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

- که در آن $\mathbf{y}_2$ ماتریسی از متغیرهای توضیحی درونزا است که در برگیرنده g متغیر و ماتریس $\mathbf{x}_1$ نیز دارای K متغیر برونزاست.

- ماتریس $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$ مجموعه‌ای از متغیرهای ابزاری است.
- و بردار اجزاء اخلال مدل داده‌های پانل یک عاملی One-Way اثرات ثابت

$$\mathbf{u} = \mathbf{Z}_\mu . \boldsymbol{\mu} + \mathbf{v}$$

برآورد مدل اثرات ثابت و متغير توضيحي درونزا

- نخست لازم است تمام متغیرهای مدل پانل اثرات ثابت میانگین‌زدایی بشوند.
- یعنی، که در آن $\tilde{\mathbf{y}} = \mathbf{Qy}$ ، $\tilde{\mathbf{Z}} = \mathbf{QZ}$ ، $\tilde{\mathbf{X}} = \mathbf{QX}$ و $\tilde{\mathbf{u}} = \mathbf{Qu}$ است.
- آنگاه:

$$\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \tilde{\mathbf{X}}'\tilde{\mathbf{Z}}\boldsymbol{\delta} + \tilde{\mathbf{X}}'\tilde{\mathbf{u}} \qquad (۱۲)$$

- آیا $\tilde{\mathbf{X}}'\tilde{\mathbf{u}}$ فرضهای کلاسیک را برقرار می‌سازد؟ خیر
- اگرچه در این ماتریس جزء اخلال جدید $\tilde{\mathbf{X}}$ مستقل از $\tilde{\mathbf{u}}$ است، ولی $\tilde{\mathbf{X}}'\tilde{\mathbf{u}}$ بعضی از فرضهای کلاسیک را برقرار نمی‌سازد.

$$E[\tilde{\mathbf{X}}'\tilde{\mathbf{u}}] = \tilde{\mathbf{X}}'\mathbf{Q}E[\mathbf{u}] = \mathbf{0}$$

$$E[\tilde{\mathbf{X}}'\tilde{\mathbf{u}}\tilde{\mathbf{u}}'\tilde{\mathbf{X}}] = [\tilde{\mathbf{X}}'\mathbf{Q}E(\mathbf{uu}')\mathbf{Q}'\tilde{\mathbf{X}}] = \sigma_v^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}) \neq \sigma_v^2\mathbf{I}_{N.T}$$

- این ماتریس قطری نیست، در نتیجه لازم است با استفاده از روش GLS با معکوس $\sigma_v^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ برآوردها صورت گیرند.

$$\hat{\boldsymbol{\delta}}_{w2sls} = (\tilde{\mathbf{Z}}'\underbrace{\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'}_{\mathbf{P}_{\tilde{\mathbf{X}}}}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\underbrace{\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'}_{\mathbf{P}_{\tilde{\mathbf{X}}}}\tilde{\mathbf{y}}$$

برآورد مدل اثرات تصادفي و متغير توضيحي درونزا

- مدل اثرات تصادفی و ماتریس واریانس کواریانس جزء اخلال یک عاملی با اثرات فردی آن یعنی $\mathbf{\Omega}$ را در نظر بگیرید.
- با پیش ضرب کردن $\mathbf{\Omega}^{-1/2}$ (تعریف شده و ...) در رگرسیون پانل رابطه وزنی شده زیر را بدست می‌آوریم.

$$\mathbf{\Omega}^{-1/2}\mathbf{y} = \mathbf{\Omega}^{-1/2}\mathbf{Z}\boldsymbol{\delta} + \mathbf{\Omega}^{-1/2}\mathbf{u}$$

- اکنون دو مشکل وجود دارد. تصادفی بودن متغیرهای توضیحی $\mathbf{Y}_2$ و برقرار نشدن فرضهای کلاسیک رگرسیون در معادله رگرسیون جدید. با قرار دادن $\mathbf{y}^* = \mathbf{\Omega}^{-1/2}\mathbf{y}$ ، $\mathbf{Z}^* = \mathbf{\Omega}^{-1/2}\mathbf{Z}$ و $\mathbf{u}^* = \mathbf{\Omega}^{-1/2}\mathbf{u}$ داریم.

$$\mathbf{y}^* = \mathbf{Z}^*\boldsymbol{\delta} + \mathbf{u}^*$$

- امید ریاضی $\mathbf{u}^* = \mathbf{\Omega}^{-1/2}\mathbf{u}$ برابر با بردار صفر و ماتریس واریانس و کواریانس آن به صورت:

$$E[\mathbf{u}^*\mathbf{u}'^*] = [\mathbf{\Omega}^{-1/2}E(\mathbf{u}\mathbf{u}')\mathbf{\Omega}^{-1/2}] = \mathbf{\Omega}^{-1/2}\mathbf{\Omega}\mathbf{\Omega}^{-1/2}$$

- می‌توان به سادگی نشان داد که $\mathbf{\Omega}^{-1/2}\mathbf{\Omega}\mathbf{\Omega}^{-1/2} = \mathbf{I}_{N.T}$ است.
- ملاحظه می‌شود که $\mathbf{u}^*$ تمام فرض‌های CLR را به غیر $E[\mathbf{u}^* \mid \mathbf{Z}^*] = \mathbf{0}$ را برقرار می‌سازد.

برآورد مدل اثرات تصادفي و متغير توضيحي درونزا

- اکنون این پرسش مطرح می‌شود که مشکل درونزایی متغیرهای ماتریس $\mathbf{Y}_2$ را چگونه می‌توان حل کرد، به گونه‌ای که تخمین‌زن‌های بدست آمده برای $\boldsymbol{\delta}$ سازگار باشند.

- برای حل مشکلات می‌توان روش IV را با متغیرهای ابزاری $\mathbf{X}^* = \mathbf{\Omega}^{-1/2}\mathbf{X}$ با $\mathbf{X} = [\mathbf{X}_1 \vdots \mathbf{X}_2]$ بکار بست.
- اکنون در نظر داشته باشید که در

$$\mathbf{X}'^*\mathbf{y}^* = \mathbf{X}'^*\mathbf{Z}^*\boldsymbol{\delta} + \mathbf{X}'^*\mathbf{u}^*$$

- تبدیل شده، اگر چه عناصر $\mathbf{u}^*$ ، نا بسته هم توزیع است، ولی دیگر $\mathbf{X}'^*\mathbf{u}^*$ این ویژگی را ندارد.
- برتری‌های معادلهٔ بالا در این است که مشکل درونزایی $\mathbf{Y}_2$ را حل می‌کند، با این حال می‌توان بکارگیری GLS برای معادله بالا ضرایب $\boldsymbol{\delta}$ برآورد نمود.
- زیرا $E[\mathbf{X}'^*\mathbf{u}^*\mathbf{u}'^*\mathbf{X}^*] = (\mathbf{X}'^*\mathbf{X}^*)$ است. تخمین‌زن‌های اثرات تصادفی $\boldsymbol{\delta}$ به صورت:

$$\hat{\boldsymbol{\delta}}_{RE2sls} = (\mathbf{Z}'^*\mathbf{P}_{\mathbf{X}^*}\mathbf{Z}^*)^{-1}\mathbf{Z}'^*\mathbf{P}_{\mathbf{X}^*}\mathbf{y}^*$$

- اندیس 2SLS به این خاطر است که در محاسبه $\mathbf{X}^*$ نیاز به در اختیار داشتن $\mathbf{\Omega}^{-1/2}$ داریم که ابتدا باید محاسبه گردد.

برآورد مدل اثرات تصادفي و متغير توضيحي درونزا

```
webuse nlswork, clear
 * Fixed-effects model
 xtivreg ln_w  c.age##c.age not_smsa (tenure = union s
 outreg2 using between, tex replace sideway
* GLS random-effects model
 xtivreg ln_w  c.age##c.age not_smsa i.race (tenure =
 outreg2 using between, tex append sideway
```

برآورد مدل اثرات تصادفي و متغير توضيحي درونزا

VARIABLES	(FE) ln_wage coef	(2) se	(RE) ln_wage coef	(4) se
ln_wage				
tenure	***0.240	(0.0373)	***0.139	(0.00786)
age	0.0118	(0.00900)	***0.0279	(0.00541)
c.age#c.age	***0.00121_	(0.00019)	***0.00083_	(8.70e-05)
not_smsa	0.0167_	(0.0339)	***0.223_	(0.0111)
2.race			***0.206_	(0.0126)
3.race			**0.110	(0.0517)
Constant	***1.678	(0.163)	***1.336	(0.0844)
Observations	19,007		19,007	
Number of id	4,134		4,134	

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

آماره دیویدسون‌ـ مکینون برای فرضیه برونزایی متغیرهای توضیحی

- ویژگی های متغیرهای ابزاری این است که این متغیرها باید ناهمبسته با جزء اخلال مدل مورد مطالعه بوده و همبستگی نزدیک با متغیر توضیحی مشکوک به درونزایی باشد.
- دیویدسون و مکنیون (1993) یک روش اقتصادسنجی ساده برای آزمودن این فرضیه ابداع کردند.
- این توابع نمونه ای به صورت آماره χ^2_g یا F است.
- برتری این روش این است که هم می توان برونزایی تک تک متغیرهای توضیحی مشکوک به درونزایی را آزمون کرد و هم می توان برونزایی تمام یا بخشی از ان را آزمون فرضیه نمود.
- ایده ی پایه ای این تکنیک آماری بسیار ساده است.

آماره دیویدسون ـ مکینون برای فرضیه برونزایی متغیرهای توضیحی

الف) فرم تحویل یافته را برای بردار $\mathbf{y}_{it2}$ با برازش هر یک از مولفه‌های $\mathbf{y}_{it2}$ بر روی تمام متغیرهای برونزای موجود (شامل توضیحی و ابزاری) برآورد و بردار $\mathbf{V}_{2it}$ برای $i=1,2,..,N$ و $t=1,2,..,T$ محاسبه نمایید.

ب) بردار $\hat{\mathbf{V}}_{2it}$ رابه عنوان متغیرهای توضیحی به معادله

$$y_{it}=\mathbf{x}'_{1it}\boldsymbol{\beta}+\mathbf{y}'_{2it}\boldsymbol{\gamma}+u_{it}$$

اضافه نمایید. این معادله کامل شده را بوسیله OLS برآورد نمایید.

- اگر ضریب این متغیرهای $\hat{\mathbf{V}}_{2it}$ از نظر آماری اختلاف معنی‌داری از صفر داشتند، آنگاه نتیجه بگیرید که $\mathbf{y}_{it2}$ درونزا است.

- در محاسبات تجربی فایل $dmexogxt$ برای اجرای این آزمون فرضیه طراحی شده است که پس از اجرای، $xtivrey$ می‌تواند اجرا شود.

آماره دیویدسون‌ـ‌مکینون برای فرضیه برونزایی متغیرهای توضیحی

- شناسایی درونزايي متغيرهاي توضيحي در يك مدل داده‌هاي پانل با اثرات ثابت

```
ssc install dmexogxt
webuse nlswork , clear
* Fixed-effects model
gen age2=age^2
xtivreg ln_w  age age2 not_smsa (tenure = union south
 dmexogxt     tenure
* GLS random-effects model
xtivreg ln_w  age age2 not_smsa i.race (tenure = unio
```

```
F  test that all u_i=0:     F(4133,14869) =     1.44      Prob > F     = 0.0000

Instrumented:   tenure
Instruments:    age age2 not_smsa union south


    dmexogxt    tenure

Davidson-MacKinnon test of exogeneity:  215.8534  F( 1,14868)  P-value =  1.6e-48
```