

وزارة التعليم العالي والبحث العلمي  
جامعة محمد بوضياف بالمسيلة  
كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قسم العلوم الاقتصادية

# الاقتصاد القياسي لبيانات البائل

محاضرات مع أمثلة

مطبوعة غير محكمة - نسخة 1.

صالح بوعبد الله

أفريل 2020



## المحتويات

3.....	المحتويات
5.....	مخطط المقياس
6.....	البرنامج
8.....	فصل 1. البيانات المقطعية المجمعة
8.....	1. مفاهيم أساسية
8.....	1-1. البيانات المقطعية المجمعة (المستقلة)
9.....	2-1. بيانات البائل
11.....	3-1. ترميز المتغيرة المفسرة الإسمية بمتغيرات ثنائيّة
14.....	2. المعالجة الإحصائية لنموذج البيانات المجمعة
14.....	1-2. السماح بتغير الثابت (أثر الزمن)
16.....	2-2. تفاعل المتغيرات المفسرة مع الزمن (تغير أثر العوامل عبر الزمن)
19.....	3-2. فرق الفروق (أثر السياسة أو الحدث)
26.....	فصل 2. نموذج الأثر الفردي وتقديره بطريقة POLS وطريقة FD
26.....	1. نموذج الأثر الفردي وتقديره بطريقة POLS
26.....	1-1. مبرر بيانات البائل
27.....	1-2. نموذج الأثر الفردي Individual-specific effects model
29.....	3-1. طريقة POLS
31.....	2. دالة الفروق الأولى
31.....	1-2. تعريف بالطريقة
32.....	2-2. شروط دالة الفروق الأولى
33.....	3-2. عوائق طريقة الفروق الأولى
34.....	4-2. حالة تعدد المتغيرات المفسرة
35.....	2-5. النماذج ذات التأخرات
36.....	6-2. تطبيق دالة الفروق الأولى على تقييم السياسات
38.....	7-2. دالة الفروق الأولى لأكثر من فترتين
42.....	فصل 3. طريقتي الأثر الثابت والأثر العشوائي
42.....	1. طريقة الأثر الثابت (Fixed Effects) بالتحويل الداخلي within
42.....	1-1. الدالة
43.....	1-2. المقارنة بين طريقة FD و FE
44.....	2. طريقة الأثر العشوائي
44.....	1-2. دالة الأثر العشوائي
45.....	2-2. اختبار هاوسمن
46.....	المقارنة بين FE و RE



## مخطط المقياس

### تعريف بالمقياس

بيانات البانل هي بيانات بعدين، البعد الفردي (أفراد، مؤسسات، بلدان، ...) والبعد الزمني (سنوات، ثلاثيات، أشهر، ...). يمكن أن تكون المفردات التي يتم قياس متغيراتها هي نفسها في كل فترة (مثلا نأخذ بيانات الفقر والبطالة والمستوى التعليمي لـ 20 بلد على مدى 10 سنوات متتالية) ويمكن أن تكون المفردات في كل مرة مستقلة عن الفترات الأخرى (نأخذ في كل سنة عينة جديدة من المؤسسات ونقيس في كل منها مثلا الربح وعدد العمال ونسبة القروض). تسمى الحالة الأولى **بيانات البانل**، وتسمى الحالة الثانية **البيانات المقطعية المجمعة المستقلة**، أو بيانات البانل البسيط، وسوف نتطرق في هذا المقياس لكلا الحالتين.

### متطلبات المقياس

يعتبر هذا المقياس من الفصول المتقدمة في الاقتصاد القياسي، ويأتي في العادة بعد دراسة السلاسل الزمنية. لذلك فإن من المهم أن يكون الطالب ملما بطريقة المربعات الصغرى لتحليل الانحدار البسيط والمتعدد. سيكون التركيز في هذا المقياس على الاستخدام والتفسير الاقتصادي لمخرجات البرمجيات، وعدا ذلك يفضل أن يكون للطالب إطلاع على المفاهيم الأساسية في المصفوفات، وبأحد البرمجيات الإحصائية، *SPSS* أو *Eviews*، وعلى أي حال فهناك توقيت لتدريس هذه البرمجيات.

### الهدف من المقياس

يستهدف المقياس تعليم الطالب استخدام تقنيات بيانات البانل لتقدير نموذج تحليل انحدار متعدد يتضمن البعد الزمني مع البعد الفردي، وذلك بهدف تحليل أثر سياسة عمومية ما (مالية أو نقدية أو اقتصادية ... ) أو أحداث ما على متغيرات اقتصادية كمية.

### الكفاءات المستهدفة

- معرفة الحالات التي تستخدم فيها تقنيات بيانات البانل
- معرفة كيفية تفسير المخرجات الأساسية للبرامج المذكورة.
- معرفة الشروط التي يتطلبها استخدام نماذج البانل.

### تنظيم المقياس:

يحتوي المقياس على محاضرة وتطبيق أسبوعيا، الحجم الساعي المتاح: حوالي 12 أسبوع، الحجم الساعي المطلوب للمقياس....

<b>البرنامج</b>	
أسبوع	المضمون
	<b>ف 1 - البيانات المقطعية المجمعة (البائل البسيط)</b>
<b>2</b>	مفاهيم أساسية
	البيانات المقطعية المجمعة
	بيانات البائل
<b>3</b>	ترميز المتغيرة الإسمية
	المعالجة الإحصائية
<b>4</b>	تغير الثابت (أثر الزمن)
<b>5</b>	تفاعل العوامل مع الزمن (تغير الأثر مع الزمن)
<b>6</b>	أثر السياسة أو الحدث (فرق الفروق)
	<b>ف 2 - نموذج الأثر الثابت وتقديره بطريقة الفروق الأولى</b>
<b>7-8</b>	طريقة <i>Pooled OLS</i>
<b>9</b>	طريقة الفروق الأولى <i>FD</i>
	<b>ف 3 - نموذج الأثر الثابت ونموذج الأثر العشوائي</b>
<b>10</b>	تقدير نموذج الأثر الثابت
<b>11</b>	تقدير نموذج الأثر العشوائي
	<b>ف 4. نموذج متقدمة لبيانات البائل</b>
<b>12</b>	نموذج الأثر العشوائي المرتبط
<b>13</b>	تطبيق تقنيات بيانات البائل على هياكل أخرى للبيانات

## المراجع الأساسية:

- Wooldridge Jeffrey M. Econometric Analysis of Cross Section and Panel Data, The MIT Press Cambridge, Massachusetts London, England.2002.
- Wooldridge Jeffrey M., Introduction à l'économétrie, une approche moderne, de boeck, Belgique, 2015
- William Green, Econométrie, 5me éd. Pearson education INC., 2005, New Jersey, USA.
- Bourbonnais Régis, l'économétrie, édition Dunod, Paris, 2017.

## مواقع على الانترنت

- موقع أكاديمية الاقتصاد القياسي

*Econometrics academy*

<https://sites.google.com/site/econometricsacademy/econometrics-models/panel-data-models>

- موقع Eviews

*EViews Help\_ Structuring a Panel Workfile*

[https://www.eviews.com/help/helpintro.html#page/content%2Fpanelwf-Structuring\\_a\\_Panel\\_Workfile.html](https://www.eviews.com/help/helpintro.html#page/content%2Fpanelwf-Structuring_a_Panel_Workfile.html)

- موقع لشرح بيانات نماذج بيانات البانل بالصوت والصورة على يوتيوب:

<https://www.youtube.com/watch?v=4d0ZFHvVks4>

<https://www.youtube.com/watch?v=1SchyQ77Vfg>

- موقع (LIBSTAT) لدراسات حالة:

<http://lib.stat.cmu.edu/DASL/allmethods.html>

- موقع STEP (STatistical Education Through Problem Solving):

<http://www.stats.gla.ac.uk/steps/glossary/index.html>

- موقع متنوع من الموارد في الإحصاء والاقتصاد القياسي وأشياء أخرى عديدة للأستاذ

Jean-Pierre Cabannes <http://cabannes.org/#5>

- موقع مماثل غني بالموارد في الاقتصاد القياسي

<http://www.economicsnetwork.ac.uk/teaching/exams/econometrics.htm>

- موقع Wikipédia

11-02-2020

## فصل 1. البيانات المقطعية المجمعة

مفاهيم أساسية في البيانات المجمعة والمتغيرة الوهمية – المعالجة الإحصائية بتغير الثابت أو بالتفاعل و فرق الفروق

18-02-2020

توطئة.

ندرس في هذا الفصل نوعا من البيانات يدمج بين البيانات "العادية" أو "المقطعية"، وهي بيانات مأخوذة في لحظة زمنية ما، وبين السلاسل الزمنية التي تستمر مع الزمن. تستخدم هذه المزوجة بين بعدين في الكثير من الدراسات التي تعني بالظواهر الاقتصادية وسيورتها لقياس واختبار تطور ظاهرة ما مع الزمن أو تغير تأثيرها بعوامل معينة أو أيضا قياس واختبار أثر سياسة عمومية ما أو حدث اقتصادي معين. قسمنا هذا الفصل إلى مبحثين: نتطرق في الأول لمفاهيم أساسية عن البيانات المقطعية وبيانات البائل والمتغيرة الوهمية، وفي المبحث الثاني نتطرق للمعالجة الإحصائية لنموذج البيانات المجمعة من خلال طريقة السماح بتغير الثابت، طريقة تفاعل المتغيرة المفسرة مع الزمن وطريقة فرق الفروق.

الفي فقرات هذا الدرس سنتعرض لأمثلة عن كيفية استخدام طرق تحليل الانحدار لدراسة هذا النوع من المسائل. الأمثلة والشروح مستقاة أساسا من كتاب وولدرج<sup>1</sup> Wooldridge, 2015.

### 1. مفاهيم أساسية

تعريف البيانات المقطعية المستقلة المجمعة  
تعريف بيانات البائل  
ترميز المتغيرة الإسمية

#### 1-1. البيانات المقطعية المجمعة (المستقلة)

يقصد بالبيانات المقطعية (time) تلك البيانات المأخوذة مرة واحدة، في لحظة زمنية واحدة، فهي لا تتضمن بعد الزمن. يقصد بالبيانات المقطعية المجمعة المستقلة<sup>2</sup> (pooled cross sections) تلك البيانات المأخوذة كعينات عشوائية من مجتمع كبير في فترات مختلفة، غالبا سنوات، مع استقلال بيانات كل عينة (سنة) عن بيانات العينات الأخرى. في هذه الحالة فإن التوزيع الاحتمالي للبيانات يختلف من فترة لأخرى، مثلا توزيع الأجر يختلف من فترة لأخرى وكذلك المؤهل

<sup>1</sup> Wooldridge Jeffrey M., Introduction à l'économétrie, une approche moderne, de boeck, Belgique, 2015

<sup>2</sup> Données en coupes transversales empilées indépendantes.



العلمي، وهكذا. لأخذ هذا الاختلاف في الحسبان في نموذج تحليل الانحدار المتعدد، يمكن إما أن نسمح للثابت بأن يتغير من فترة لأخرى، أو تتغير معاملات الانحدار الجزئية، أو كلاهما.

أمثلة:

- بيانات تعداد السكان: نستخرج في 2010 عينة من 300 عائلة، ونأخذ بياناتها للدخل وعدد الأولاد، وعدد غرف السكن، المستوى الدراسي للأم. في 2020 نستخرج مجددا عينة من 320 عائلة أخرى، ونأخذ ذات البيانات المذكورة. نقوم بتجميع البيانات في جدول بيانات واحد، نقول عنها أنها بيانات مقطعية مجمعة (مستقلة). حجم العينة الإجمالي هو 620 عائلة.

Obs	Year	Income	kids	rooms	mothlevel
...	...	...	...	...	...
<b>298</b>	<b>2010</b>	<b>55000</b>	<b>2</b>	<b>3</b>	<b>8</b>
<b>299</b>	<b>2010</b>	<b>28000</b>	<b>3</b>	<b>4</b>	<b>12</b>
<b>300</b>	<b>2010</b>	<b>32000</b>	<b>4</b>	<b>4</b>	<b>6</b>
301	2020	31000	2	2	14
302	2020	29000	4	5	11
303	2020	28000	5	6	12
...	...	...	...	...	...
620	2020	46000	6	1	8

- نأخذ بيانات: 'الأجر' و'عدد ساعات العمل' و'عدد العطل المرضية' و'المؤهل المهني' لعينة من العمال في سنة 2005، ثم بعد 5 سنوات نكرر العملية على عينة جديدة من العمال، لدراسة تطور تأثير الأجر بالمتغيرات المستقلة المذكورة. يمكن أن نكرر العملية أكثر من مرة.

- نكون عينة من المؤسسات من قطاع معين في 2020 ونسجل بيانات 'الربح' لقياس تأثيره ب'رقم الأعمال' و'الإفناق الإشهاري' وبعد سنتين نكرر العملية مع عينة جديدة. الهدف هو قياس تأثير علاقة 'الربح' بالمتغيرتين المذكورتين بعد إدخال ضريبة جديدة مثلا.

## 2-1. بيانات البائل

على غرار البيانات المقطعية المجمعة، تزوج بيانات البائل بين البعدين: الفردي والزمني، لكن العينات ليست مستقلة، وذلك لأن المفردات تبقى نفسها. في بيانات البائل (تسمى أيضا البيانات الطولية Longitudinal) نتتبع نفس المفردات (مؤسسات، بلدان، مدن، عائلات، مواطنون...) لنستقي بياناتها كل فترة. العينة المكررة تتقضى الاستقلالية؛ فأجر عامل ما في سنة ما لن يكون مستقلا عن أجره في السنة السابقة، وأداء مؤسسة في سنة ما ليس مستقلا عن أدائها قبل سنتين، إلخ.

مثال: ليكن لدينا بيانات 'معدل الضريبة' و'البطالة' ل 90 دولة نامية، وهذه البيانات ممتدة من 2001 إلى 2005، ونريد أن نستخدم هذه البيانات لتفسير البطالة. إذا أدخلنا بيانات بلد واحد خلال الفترة فلن نعرف إن كانت العلاقة

تتعمم على البلدان الأخرى، وإذا إدخالنا بيانات البلدان في لحظة واحدة فلن نعرف إن كانت العلاقة معممة على مر الزمن. الحل هو أخذ بيانات مختلف البلدان على مدى الزمن، وهي البيانات المجمعة.

Obs	State	Year	TaxRate	Unemployment
...	...	...	...	...
80	Gabon	2001	.35	0.23
81	Gabon	2002	.37	0.20
82	Gabon	2003	.36	0.21
83	Togo	2001	.14	0.18
84	Togo	2002	.16	0.19
85	Togo	2003	.11	0.14
...	...	...	...	...

الحصول على بيانات بانل قد يكون سهلا في وحدات مستقرة، يمكن الرجوع إليها، مثل مؤسسات، مدارس، بلدان...، على عكس الأفراد، فالفرد قد لا يمكن الرجوع إليه بسبب تغيير العنوان أو الموت أو غيره.

**مثال 2.** نهتم بتأثير الدخل الوطني، ولدينا بيانات عينة من البلدان تم تتبعها على عدة سنوات.

**مثال 3.** نريد اختبار تأثير مستوى التعليم على دخل الفرد، ولدينا بيانات عينة من الأفراد تم تتبعهم على عدة سنوات.

**خلاصة:** استخدام بيانات البانل بالمزاوجة بين البعدين الزمني والفرد يحسن تحليل السياسات الاقتصادية، لأنه يسمح بعزل تأثير عوامل خفية عديدة لم يتم قياسها لاستحالة أو صعوبة القياس، مثل العوامل الثقافية والاجتماعية وغيرها<sup>1</sup>.  
بيانات البانل يمكن أن تكون "متوازنة" (balanced) إذا كانت كل المفردات لدينا مشاهداتها في كل الفترات ( $T_i = T$  for all  $i$ )، أو تكون "غير متوازنة" (unbalanced) إذا كانت هناك قيم مفقودة بحيث أن بعض المفردات ليس لدينا مشاهداتها لكل الفترات ( $T_i \neq T$ ).

في بيانات البانل يمكن أن يكون هناك ارتباط مشاهدات الفرد الواحد عبر الزمن، لكن ليس هناك ارتباط بين الأفراد. مثلا: دخل الفرد مرتبط ذاتيا عبر الزمن، لكن دخول الأفراد المختلفين مستقلة عن بعضها البعض.

بيانات البانل أنواع: يمكن أن تكون قصيرة (short panel data) بمعنى بيانات لمفردات كثيرة على فترات زمنية قليلة، وهي الحالة الغالبة. يمكن أن تكون طويلة (long) بمعنى فترات كيلة زمنيا ومفردات قليلة، وهي حالة قليلا ما نجدها في الدراسات الاقتصادية. والحالة الثالثة تدعى (both) وهي بيانات كثير من الأفراد على فترة طويلة، وهي حالة نادرة.

المتغيرات المفسرة للنموذج أنواع: متغيرات متغيرة ( $\text{varying regressors } x_{it}$ )، مثلا الدخل، عدد أولاد العائلة، الحالة العائلية... فهي متغيرة من فرد لآخر ومن فترى لأخرى. هناك المتغيرات الثابتة في الزمن (time-invariant regressors  $x_{it} = x_i$  for all  $t$ )، مثل الجنس، والعرق،... وهناك متغيرات ثابتة بين الأفراد (individual-invariant regressors  $x_{it} = x_t$  for all  $i$ ) مثل متغيرات التوجه الاقتصادية معدل البطالة أو التضخم.

<sup>1</sup> تسمى هذه الطريقة في المربعات الصغرى نموذج الأثر المثبت (fixed effects)، في هذه الحالة سيكون لدينا نموذج الأثر المثبت للأفراد ونموذج الأثر المثبت للزمن ويمكن استخدام نموذج للأثر المثبت لكل منهما. عيب هذه الطريقة أنها تحتاج إلى عدد كبير من المتغيرات الوهمية، وهذا يقلل من درجات الحرية للنموذج، وبالتالي فهي لا تصلح إلا إذا كان عدد الأنماط في المتغيرات النوعية ضئيلا. الطريقة الأخرى لمعالجة البيانات ذات البعدين (بيانات البانل) هي نموذج الأثر العشوائي (Random effects model)، وهذه قصة أخرى، نصل إليها فيما بعد.

## 3-1. ترميز المتغيرة المفسرة الإسمية بمتغيرات ثنائية

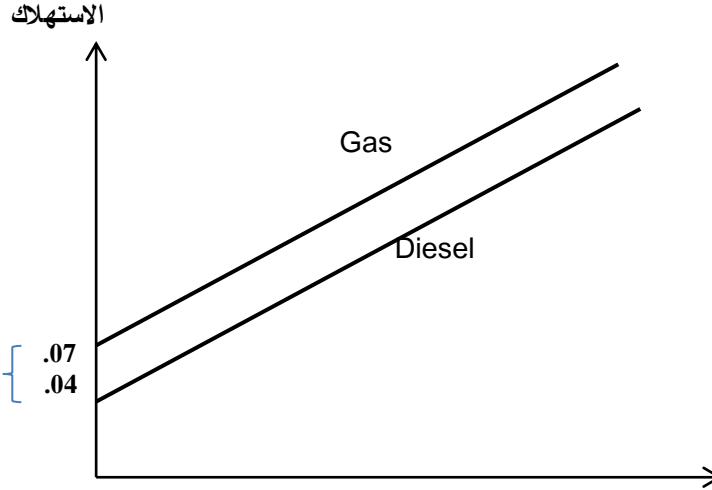
ماهي المتغيرة الإسمية وكيف ندرج متغيرات نوعية في النموذج؟ المتغيرة النوعية تأخذ أنماطاً (مثلاً متغيرة الحالة العائلية تأخذ الأنماط التالية: 'متزوج'، 'أعزب'، 'مطلق'، ...) بينما المتغيرة الكمية تأخذ قيماً عددية. نعبر عنها بمجموعة من المتغيرات الثنائية المصطنعة أو الوهمية (المتغيرة الثنائية تأخذ لكل مشاهدة إحدى القيمتين 0 أو 1: تأخذ 0 إذا لم يتحقق النمط في المشاهدة وتأخذ 1 إذا تحقق). عدد المتغيرات الوهمية (dummy variables) الضرورية لتعويض المتغيرة النوعية يساوي عدد أنماطها ناقصاً الواحد (أحد الأنماط يترك كمرجع حتى لا يتكون ارتباط مشترك تام).

**مثال 1.** نريد ربط كمية الوقود المستهلك (بالتر) بمتغيرة كمية هي المسافة المقطوعة Dist (بالكلم) ومتغيرة نوعية هي نوع الوقود Type، أنماطها "ديزل" و "بنزين". لإدراج Type نحتاج إلى متغيرة وهمية واحدة (لأن عدد الأنماط هو 2: نستخدم الرمز "0" للديزل و "1" للبنزين). نفترض أن الدالة جاءت كما يلي:

$$\text{Cons}^{\wedge} = 0.4 + 0.3 \text{ Type} + 0.6 \text{ Dist}$$

معامل الانحدار الجزئي للمتغيرة Type يمثل الفرق بين استهلاك البنزين والديزل: المركبة التي تستخدم الوقود من نوع "بنزين" (رمز النمط هو 1) تستهلك في المتوسط 0.3 لتر أكثر في الكلم الواحد مقارنة مع المركبات العاملة بالديزل، وذلك لنفس المسافة المقطوعة.

لو مثلنا الانحدار البسيط لاستهلاك الوقود على المسافة المقطوعة بمركبات البنزين لوحدها، وفعلنا الأمر نفسه مع مركبات الديزل، سيكون خط المربعات الصغرى لمركبات البنزين أعلى بمقدار 0.3 من خط مركبات الديزل.



رسم توضيحي 1. إدراج متغيرة ثنائية في النموذج. لذات المسافة يستهلك محرك البنزين وقوداً أكثر.

**مثال 2:** في المثال السابق، نفترض أن المتغيرة الإسمية Type تأخذ ثلاثة أنماط: "Gas"، "Diesel"، "Super"؛ لترميز هذه المتغيرة الإسمية نحتاج إلى:

- متغيرة D تأخذ القيمة 1 للسيارة العاملة بـ "Diesel"، و 0 إذا لم تكن كذلك،
- متغيرة G تأخذ 1 للسيارات العاملة بـ "Gas"، و 0 إذا لم تكن كذلك،

جدول 1 ترميز متغيرة نوعية ذات ثلاثة أنماط

الأنماط				
"Super"	"Diesel"	"Gas"		
0	0	1	Gas	المتغيرات
0	1	0	Diesel	

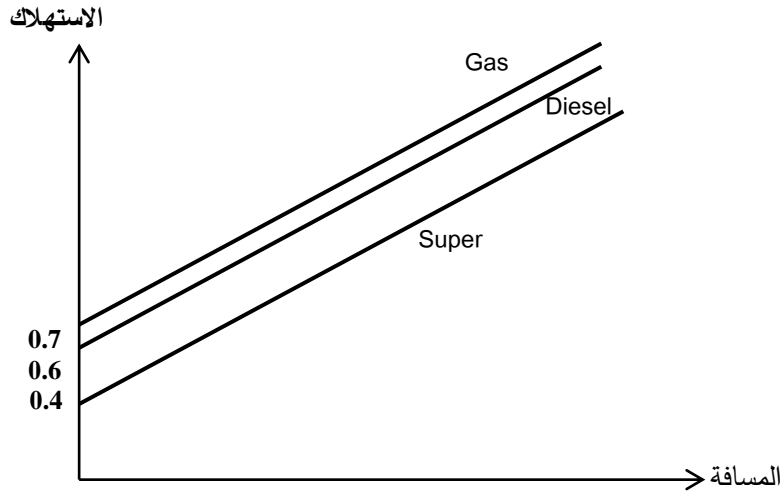
يكتب النموذج كالتالي:

$$\text{Cons} = \beta_0 + \delta_0 \text{Gas} + \delta_1 \text{Diesel} + \beta_1 \text{Distance} + \varepsilon$$

نفترض أن الدالة كانت:

$$\text{Cons}^{\wedge} = 0.4 + 0.3 \text{Gas} + 0.2 \text{Diesel} + 0.6 \text{Dist}$$

في هذه الحالة فإن معامل الانحدار الجزئي للمتغيرة Gas يمثل الفرق بين استهلاك البنزين العادي واستهلاك البنزين الممتاز، أي أنه يأخذ التفسير التالي: المركبة التي تستخدم "بنزين" (Gas) تستهلك في المتوسط 0.3 لتر أكثر في الكلم الواحد مقارنة مع المركبات العاملة بـ "الممتاز"، وذلك لنفس المسافة المقطوعة. معامل الانحدار الجزئي للمتغيرة Diesel يمثل الفرق بين استهلاك الديزل واستهلاك الممتاز، أي أنه يأخذ التفسير التالي: المركبة التي تستخدم الوقود من نوع Diesel تستهلك في المتوسط 0.2 لتر أكثر في الكلم الواحد مقارنة مع المركبات العاملة بالممتاز، وذلك لنفس المسافة المقطوعة. لو مثلنا الانحدار البسيط لاستهلاك الوقود على المسافة المقطوعة بمركبات البنزين لوحدها، وفعلنا الأمر نفسه مع مركبات الديزل والسوبر، سيكون خط المربعات الصغرى لمركبات البنزين أعلى بمقدار 0.3 من خط المربعات الصغرى لمركبات الديزل.



رسم توضيحي 2. إدراج متغيرة ثنائية في النموذج. لذات المسافة يستهلك محرك البنزين وقودا أكثر.

سؤال: قم بصياغة الدالة في حالة اتخاذ النمط Gas كمرجع.

**مثال 3<sup>1</sup>:** بفرض أننا أمام نموذج يتوقع كمية الوقود المستهلك بناء على نوع السيارة، وتأخذ أربعة أنماط: "رياضية"، "تجارية"، "سياحية"، و"عائلية". الترميز سيكون على غرار المثال السابق ولكن بثلاث متغيرات ثنائية.

جدول 2 ترميز متغيرة نوعية ذات أربعة أنماط

أنماط المتغيرة الاسمية				
"رياضية"	"تجارية"	"عائلية"	"سياحية"	
1	0	0	0	D <sub>1</sub>
0	1	0	0	D <sub>2</sub>
0	0	1	0	D <sub>3</sub>

أكتب الدالة، وأستخرج كمية الوقود المستهلك لمختلف أنواع السيارات الأربع.  
الجواب:

$$\hat{y} = b_0 + b_1D_1 + b_2D_2 + b_3D_3$$

النمط الرابع "سياحية" اتخذ مرجعا ولم يدرج في الدالة.

يظهر هذا المثال العلاقة بين تحليل الانحدار البسيط بمتغيرة إسمية وتحليل التباين الأحادي. التباين المتبقي هو نفسه في النموذجين، والتباين المفسر هونفسه أيضا، وكذلك اختبار F لمجمل النموذج، أما المعاملات فتفسر كما يلي:

$$D_1 = D_2 = D_3 = 0 \quad \text{- السيارات السياحية:}$$

إذن تأخذ الدالة القيمة  $b_0$  ويمثل الاستهلاك المتوقع للسيارة السياحية:

$$\hat{y} = b_0 + b_1(0) + b_2(0) + b_3(0) = b_0$$

- للسيارات الرياضية:

$$D_1 = 1, D_2 = D_3 = 0 \Rightarrow \hat{y} = b_0 + b_1(1) + b_2(0) + b_3(0) = b_0 + b_1$$

$b_1$  يمثل الفرق المتوقع بين السيارات السياحية والسيارات الرياضية في استهلاك الوقود.

- للسيارات التجارية:

$$D_1 = 0, D_2 = 1, D_3 = 0 \Rightarrow \hat{y} = b_0 + b_2$$

المعامل  $b_2$  يفسر بالفرق المتوقع بين السيارات التجارية والسياحية.

- للسيارات العائلية:

$$D_1 = 0, D_2 = 0, D_3 = 1 \Rightarrow \hat{y} = b_0 + b_3$$

المعامل  $b_3$  يفسر بالفرق بين القيمة المتوقعة بين السيارات العائلية والسياحية.

25-02-2020

عرفنا إذن في هذا المبحث أن البيانات المقطعية المجمعة - أو بيانات البائل البسيط - هي بيانات لعينات مستقلة من المفردات تأخذ في فترات زمنية متتالية، فترتين أو أكثر، بينما بيانات البائل هي بيانات عينة من المفردات يتم تتبعها على فترات متتالية، فترتين أو أكثر. عرفنا أيضا كيفية الترميز للمتغيرة الاسمية باستخدام متغيرة وهمية ثنائية أو عدد من المتغيرات الثنائية الوهمية وتفسير معاملات هذه المتغيرات الوهمية. في المبحث الموالي نتطرق للمعالجة الإحصائية لنموذج البيانات المقطعية المجمعة وبيانات البائل، حيث نرى أن هناك أكثر من طريقة لإدراج عامل الزمن في النموذج.

## 2. المعالجة الإحصائية لنموذج البيانات المجمعة

من أسباب استخدام البيانات المقطعية المجمعة المستقلة (أو اختصارا البيانات المجمعة pooled data) الرغبة في زيادة حجم العينة. البيانات المجمعة عن نفس المجتمع على فترات تعطي مقدرات أكثر دقة واختبارات أقوى. في هذه الحالة بالذات، يكون التجميع مفيدا إذا كانت علاقة المتغيرة التابعة ثابتة عبر الزمن على الأقل ببعض المتغيرات المستقلة (وإلا إذا كانت كل العلاقات تتغير يصبح استخدام نماذج مختلفة هو الحل بدون الحاجة للتجميع).

يترتب على تجميع بيانات من فترات مختلفة بعض التعقيدات الإحصائية لكي يمكن احتساب اختلاف التوزيع الاحتمالي للمتغيرات. يمكن أن يتم ذلك إما بالسماح بتغير الثابت، أو بالسماح بتداخل الزمن مع معاملات المتغيرات المستقلة.

### 1-2. السماح بتغير الثابت (أثر الزمن)

في هذه الطريقة ندرج الزمن كمتغيرة مفسرة في النموذج ولكن باعتباره متغيرة إسمية وليس كمية. يكون ذلك بإدراج متغيرة مفسرة ثنائية لكل سنة إلا سنة الأساس أو "المرجع"، وغالبا ما تكون هي السنة الأولى (لكن ليس بالضرورة). أحيانا يكون تطور المعاملات السنوية في حد ذاته مهما.

النموذج الذي يمكن من احتساب أثر الزمن (بفرض أن لدينا بيانات K فترة و J عامل) يكون كالتالي:

$$y = \beta_0 + \delta_0 \text{year}_2 + \delta_1 \text{year}_3 + \dots + \delta_{K-2} \text{year}_K + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_J x_J + e$$

المعامل  $\beta_0$  هو الثابت (في حالة السنة 1).

$\beta_0 + \delta_0$  يمثل الثابت بالنسبة لمشاهدات السنة الثانية (السنوات الأخرى تختفي لأن معاملاتها تتعدم، وتبقى العوامل  $X_J$ )

$\beta_0 + \delta_1$  يمثل الثابت بالنسبة لمشاهدات السنة الثالثة (السنوات الأخرى تختفي لأن معاملاتها تتعدم، وتبقى العوامل  $X_J$ )

...

المعامل  $\beta_1$  يقيس الأثر الجزئي للعامل الأول  $X_1$  عند تثبيت العوامل الأخرى

...

$\delta_0$  يمثل التغير في التابع من السنة الأولى (سنة المرجع) إلى السنة الثانية عند تثبيت العوامل الأخرى

$\delta_1$  يقيس التغير في التابع من السنة الأولى إلى السنة الثالثة عند تثبيت العوامل الأخرى

...

المعامل  $\beta_J$  يقيس الأثر الجزئي للعامل  $X_J$  عند تثبيت العوامل الأخرى

مثلا<sup>1</sup> (الخصوبة): في "دراسة عامة اجتماعية" قام بها "مركز البحث في الرأي الوطني" في الولايات المتحدة، عولجت المسألة التالية: هل اختلفت الخصوبة (kids) لدى النساء الأكثر من 35 سنة بين سنة 1972 وسنة

<sup>1</sup> Wooldridge J. M., p. 644.

1984؟ وهذا بعد احتساب العوامل الديمغرافية التالية: سنوات التعليم educ، العمر age، العمر عند الزواج age2، العرق Black، المدينة والجهة، وعوامل أخرى. كيف يكون النموذج وتفسيره؟ المعاينة تمت كل سنتين.

جدول 3 جدول بيانات مقطعية مجمعة (pooled data). هنا الصورة وهمية لتوضيح شكل جدول البيانات.

Obs	individuals	Y74	Y76	Y78	Y80	Y82	Y84	kids	educ	age	age2	black	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
199	Pamila	0	0	0	0	0	0	3	10	36	30	0	...
200	Linda	0	0	0	0	0	0	4	12	36	22	1	...
201	Sofia	1	0	0	0	0	0	2	11	37	18	1	...
202	Lilia	1	0	0	0	0	0	1	9	35	28	0	...
203	Kati	1	0	0	0	0	0	0	8	35	38	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
401	Salv	0	1	0	0	0	0	5	5	40	41	1	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1129	Margerit	0	0	0	0	0	1	1	14	42	39	0	...

Xj	Coefficients	SE
Educ	-0.128	0.018
Age	0.532	0.138
Age2	-0.0058	0.0016
Black	1.076	0.174
East	0.217	0.133
Northen	0.363	0.121
West	0.198	0.167
Farm	-0.053	0.147
Othrural	-0.163	0.175
Town	0.084	0.124
Smcity	0.212	0.160
<b>Y74</b>	0.268	0.173
<b>Y76</b>	-0.097	0.179
<b>Y78</b>	-0.069	0.182
<b>Y80</b>	-0.071	0.183
<b>Y82</b>	-0.522	0.172
<b>Y84</b>	-0.545	0.175
<b>Constant</b>	-7.742	3.052
N = 1129, R <sup>2</sup> = 0.1295, R <sup>2</sup> aj = 0.1162		

- ما هو مقدار التغير في الخصوبة (عدد الأولاد) للمرأة من سنة 72 إلى 74، من 72 إلى 76، من 72 إلى 78، من 72 إلى 80، من 72 إلى 82، من 72 إلى 84، عند تثبيت العوامل الأخرى؟
- ج: من 72 إلى 84 انخفضت الخصوبة في المتوسط ب 0.545، أي في المتوسط، لكل مئة امرأة 55 طفل أقل.

- هل أسباب تغير الخصوبة المحسوبة بمعاملات Y74 إلى Y84 هي المتغيرات المذكورة قبل في الجدول؟  
ج: لا. المعاملات تحتسب الأثر الجزئي، أي مع تثبيت العوامل الأخرى.
- س: هل النساء الذين زاولن التعليم لمدة أطول كان لهن أولاد أقل؟ ما هو مقدار الفرق في الخصوبة بين ذوات التعليم الثانوي وذوات التعليم الجامعي (3 سنوات فرق)، عند تثبيت العوامل الأخرى؟  
الجواب هو نعم، في ثلاث سنوات انخفضت الخصوبة بمعدل  $(0.128 \times 3 = 0.384)$ .
- هل النساء ذوات العرق الأسود هن أكثر خصوبة، بكم؟ ج: نعم، بطفل تقريبا.
- علق على قيمة  $R^2 = 0.1295$ . ج: قيمة ضئيلة، وهذا لأن عدد الأطفال تآثر عليه عوامل كثيرة أخرى، لم يتم احتسابها ويصعب احتسابها جميعا، ولذلك ليس من النادر رؤية قيمة ضئيلة ل  $R^2$  في الدراسات الاقتصادية والاجتماعية والإنسانية عموما.

رأينا في المثال أعلاه أن نموذج تغير الثابت يعطينا إذن تطور المتغيرة التابعة مع الزمن، فهذا هو الهدف الأساسي لهذه الطريقة، لكنه لا يبين تغير العوامل مع الزمن، لأنه في الواقع يفترض أن العوامل الأخرى ثابتة مع الزمن. في الفقرة الموالية نتطرق لحالة تغير أثر العوامل عبر الزمن وكيفية حسابه.

## 2-2. تفاعل المتغيرات المفسرة مع الزمن (تغير أثر العوامل عبر الزمن)

في المثال السابق استخدم تغير الثابت لقياس أثر الزمن على المتغيرة التابعة. يمكن أيضا قياس تغير الأثر نفسه مع الزمن، أي تغير أثر المغيرة أو المتغيرات المفسرة على التابعة مع الزمن. يكون ذلك بإدراج تفاعل معاملات المتغيرات المفسرة مع الزمن.

بفرض أن لدينا بيانات K فترة و J عامل، يمكن أن نكتب النموذج كالتالي:

$$y = \beta_0 + \delta_0 \text{year}_2 + \delta_1 \text{year}_3 + \dots + \delta_{k-2} \text{year}_k + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \delta_k \text{year}_k x_j + \dots + e$$

المعامل  $\beta_0$  هو الثابت بالنسبة لمشاهدات سنة الأساس.

$\beta_0 + \delta_0$  هو الثابت بالنسبة لمشاهدات السنة الثانية (لأن السنوات الأخرى معدومة).

$\delta_0$ : يقيس التغير في التابع من السنة الأولى إلى السنة الثانية عند تثبيت العوامل الأخرى

$\delta_1$ : يقيس التغير في التابع من السنة الأولى إلى السنة الثالثة عند تثبيت العوامل الأخرى

...

$\delta_k$  في الحد  $\delta_k \text{year}_k x_j$  يمثل تغير أثر العامل  $x_j$  بين سنة الأساس وسنة المقارنة k.

يمكن بهذه الطريقة قياس تغير أثر العامل أو العوامل  $x_j$  مع الزمن. عمليا يتطلب هذا إنشاء متغيرة جديدة هي  $\text{year}_k x_j$ .



مثال<sup>1</sup> 2 (الأجر): في دراسة على تغير تأثير كل من عدد سنوات التعليم educ، والنوع female، على الأجر wage بين سنتي 1978 و1985، مع تثبيت أثر عدد سنوات الخبرة exper والانتماء إلى النقابة union جاء، النموذج كما يلي:

$$\text{Log(wage)} = 0.459 + 0.118 \text{ y85} + 0.0747 \text{ educ} + 0.0185 \text{ y85.educ} + 0.0296 \text{ exper} - 0.0004 \text{ exper}^2 + 0.202 \text{ union} - 0.317 \text{ female} + 0.085 \text{ y85.female} + e$$

$$\text{SE} = \begin{matrix} (0.093) & (0.124) & (0.0067) & (0.0094) & (0.0036) \\ (0.00008) & (0.03) & (0.037) & (0.051) & \end{matrix}$$

$$N = 1084; R^2 = 0.426; R^2_{aj} = 0.422$$

y85 هي متغيرة ثنائية تأخذ 1 إذا كانت المشاهدة مسجلة في عينة 85، وتأخذ 0 بخلاف ذلك. Union تأخذ 1 إذا كان العامل ينتمي لنقابة، وتأخذ 0 بخلاف ذلك و female تأخذ 1 للنساء و 0 للرجال.

• ما هو الثابت بالنسبة لمشاهدات 78؟ الجواب: 0.459

• ما هو الثابت بالنسبة لمشاهدات 85؟ الجواب: 0.459 + 0.118

• ماذا تمثل 0.118 + ؟ ج: تمثل زيادة أو توجه عام زائد في لغ الأجر من 78 إلى 85 لفائدة الجميع بصرف النظر عن الجنس والدراسة والخبرة والانتماء للنقابة...

• ما هو مردود التعليم (أثر سنة تعليم إضافية على الأجر) في 78؟ الجواب: هو معامل educ أي: 0.0747

• ما هو مردود التعليم في سنة 85؟ الجواب: 0.0747 + 0.0185

• ماذا يقيس معامل y85.educ؟ الجواب: القيمة 0.0185 تقيس التغير في مردود سنة تعليم إضافية، بين الفترتين 78 و 85.

• ماذا يقيس معامل female؟ ج: المعامل -0.317 يقيس فارق الأجر بين الجنسين.

• ماذا يقيس معامل y85.female؟ ج: المعامل 0.085 يقيس التغير في فارق الأجر بين الجنسين، بين الفترتين 78 و 85 (كونه عكس إشارة المعامل female يعني أن الفارق تقلص).

• كيف يمكن اختبار فرضية أن فارق الأجر بين النساء والرجال لم يتغير بين 78 و 85؟

الجواب: باختبار فرضية أن معامل (y85.female) يساوي الصفر: نقسم المعامل 0.085 على الخطأ المعياري لنحصل على t ونقارنها مع القيمة الجدولية، أو مع 1.96، إذا جاءت أكبر نقول إن المعامل دال، عند مستوى معنوية 5% (اختبار ثنائي).

• ما هو تأثير اللغزتم على تفسير المعاملات؟

<sup>1</sup> Wooldridge J. M., p. 646.

الجواب: إدخال اللغزتم على التابع يجعل معاملات الانحدار الجزئية تمثل نسبا (عند ضربها في مئة تصبح نسبا مئوية، أي نسب الزيادة في التابع عند زيادة المستقل بوحدة واحدة) تقريبا. مثلا: معامل عدد سنوات التكوين educ يفسر هكذا: في المتوسط فإن سنة تكوين إضافية تؤدي إلى زيادة الأجر ب 7.47 بالمائة تقريبا، أي أن مردود سنة تكوين في 78 هو 7.5% تقريبا، مردود سنة تكوين في 85 هو 9.35% تقريبا أي (1.85 + 7.5).

• قم باختبار تغير مردود التعليم educ بين 78 و 85.

الجواب: إحصائية الاختبار t هي نسبة المعامل إلى خطئه المعياري:  $1.97 = 0.0185/0.0094$ ، لذلك فهي دالة عند مستوى معنوية 5% (اختبار ثنائي). نعم هناك تغير في مردود التكوين، وبما أن المعامل 0.0185 موجب فالتغير هو بالزيادة.

• بكم كان يختلف أجر النساء عن الرجال في 78 في المتوسط؟ الجواب: في 78 كان أجر المرأة يقل عن أجر الرجل في المتوسط بنسبة 31.7% تقريبا (التقدير الحقيقي هو  $27.2\% = 1 - \exp(-0.317)$ ).

• قم بقياس واختبار إنخفاض فارق الأجر بسبب الجنس بين الفترتين.

الجواب: في 85، أجر المرأة باللغزتم يقل عن أجر الرجل في المتوسط ب 0.232 (-0.317 + 0.085).

إحصائية الاختبار تساوي  $1.67 = 0.085/0.51$ ، فهي دالة عند مستوى معنوية 5%، ( $t_{tab} = 1.64$  في الإختبار الأحادي من اليمين، حيث الفرضية البديلة هي أن معامل التفاعل أكبر من 0).

• ماذا يحدث إذا فاعلنا كل المتغيرات المفسرة بالمتغيرة y85 في الدالة أعلاه؟

الجواب: هذا يماثل استخدام دالتين، أحدهما لسنة 78 والأخرى لسنة 85. يمكن عند ذلك المقارنة بين الدالتين، واستخدام اختبار شاو Le test de Chow لوجود اختلاف بين الدالتين. كما ذكرنا سابقا، لا يعود التجميع مفيدا إذا كانت علاقة المتغيرة التابعة ثابتة عبر الزمن على الأقل ببعض المتغيرات المستقلة، وإلا (إذا كانت كل العلاقات تتغير) يصبح استخدام نموذج لكل سنة هو الحل بدون الحاجة للتجميع (~إضافة إلى أن استخدام نماذج متعددة، واحد لكل سنة، يعني تقليص حجم العينة لكل نموذج، تماما كما أن تعدد المتغيرات بإدراج التداخلات يتطلب عينة أكبر).

• كيف يمكن احتساب التضخم في هذا النموذج (قيمة العملة بعد 7 سنوات لا تبقى على حالها، فكيف نتأكد أن التضخم لا يؤثر على المعاملات)؟

الجواب: الطريقة المباشرة هي أن نستخدم القيمة الحقيقية لا الإسمية للأجر، وذلك باحتساب بيانات أجر عينة 85 بعملة 78، وذلك عن طريق التكميش deflateur (في الدراسة كان معامل التكميش هو  $1.65 = 107.6/65.2$ )، أي قسمة كل بيانات الأجر ل 85 على المعامل 1.65. لكن هذا ليس ضروريا طالما أن لدينا المفسرة y85 (تغيير الثابت بين الفترتين) وأن المتغيرة التابعة هي لغزتم الأجر. إذا لم نستخدم تغيير الثابت فإن استخدام الأجر الإسمي يمكن أن يعطي نتائج خاطئة، وإذا لم نستخدم اللغزتم يتعين استخدام متغيرة تمثل السنة واستخدام بيانات الأجر الحقيقي لا الإسمي. التفسير هو أن استخدام الأجر الحقيقي (مكمش) لا يؤثر هنا على المعاملات وإنما فقط على الثابت:

$$\text{Log}(\text{wage}/1.65) = \text{log}(\text{wage}) - \text{log}(1.65)$$

## 3-2. فرق الفروق (أثر السياسة أو الحدث)

التفاعل بين الزمن والمتغيرة له تطبيقات عديدة، خاصة عندما تأتي البيانات من تجربة طبيعية (أو شبه طبيعية). الهدف هو قياس أثر سياسة أو حدث ما على المفردات (أفراد، مؤسسات، عائلات، مدن، ...)، وللقيام بذلك يتم تكوين عينتين، عينة مراقبة *échantillon de contrôle*، لا تخضع للحدث، ولنرمز لها بـ C، وعينة معالجة *échantillon de traitement* تخضع للحدث، نرمز لها بـ T. في الوقت نفسه، نحتاج إلى بيانات كل من العنيتين قبل وبعد المعالجة (أو الحدث أو السياسة). بهذا يكون لدينا أربع عينات: عينة المراقبة قبل؛ عينة المراقبة بعد؛ عينة المعالجة قبل؛ وعينة المعالجة بعد.

مثلاً، لاختبار تأثير عملية تأهيل المؤسسات على أدائها، نأخذ أداء عينة من المؤسسات التي خضعت للتأهيل (عينة معالجة) وعينة من المؤسسات التي لم تخضع للتأهيل (عينة مراقبة). في الوقت نفسه، نحتاج أن تقسم البيانات إلى عينتين قبل وبعد فترة التأهيل لاحتساب الفروق التي تميز العنيتين، أي قبل الحدث وبعد الحدث. في المجموع سيكون لدينا أربع مجموعات من المؤسسات: عينة المراقبة قبل؛ عينة المراقبة بعد؛ عينة المعالجة قبل؛ وعينة المعالجة بعد.

لتمييز هذه البيانات نستحدث متغيرتين:

- نستحدث متغيرة  $dT$ : تساوي 1 لمشاهدات عينة المعالجة، و 0 لغيرها،
- نستحدث متغيرة  $d2$ : تساوي 1 لمشاهدات العينة الثانية في الزمن، أي التي تلي الحدث، و 0 للأولى.

النموذج الذي يمكن من احتساب أثر الحدث على المفردات التي خضعت له يكون كالتالي (difference in differences):

$$y = \beta_0 + \delta_0 d2 + \beta_1 dT + \delta_1 d2 \cdot dT + \beta_2 x_2 + \dots + \beta_r x_r + e$$

المعامل  $\beta_0$  يقيس متوسط عينة المراقبة قبل الحدث.

$\beta_0 + \delta_0$  يعطي متوسط عينة المراقبة بعد الحدث.

المعامل  $\delta_0$  يقيس الفرق بين الفترتين لدى عينة المراقبة.

المعامل  $\beta_1$  يقيس الفرق بين عینتي المراقبة والمعالجة قبل الحدث (لماذا قبل؟ ج: تذكر أنه معامل جزئي، فهو يدرج المعالجة دون الزمن).

المعامل  $\delta_1$  يقيس أثر السياسة أو الحدث على المفردات التي خضعت له (وهذا الذي نبحث عنه، لأنه يقيس تغير الفرق بين عينة المعالجة وعينة المراقبة بعد خضوع الأولى للحدث).

في حالة عدم وجود عوامل أخرى: إذا لم يدرج في النموذج عوامل أخرى، فقط الفترة والمعالجة، فإن  $\delta_1$  يقدر فرق الفروق، أو ما يسمى بمقدر الفروق المضاعفة، ويمكن حسابه كما يلي:

$$\delta_1 = (My_{2.T} - My_{2.C}) - (My_{1.T} - My_{1.C})$$

المقدر  $\delta_1$  يسمى أحياناً متوسط أثر المعالجة (effet moyen du traitement)، هو الفرق بعد المعالجة مطروحاً منه الفرق قبل المعالجة (بين متوسطي عینتي المراقبة والمعالجة)، وهو يحسب كما يلي:

- نحسب في كل فترة الفرق بين متوسطي عيني المعالجة والمراقبة
- نحسب الفرق بين الفرقين.
- يمكن أيضا إجراء الحساب كما يلي:

$$\hat{\delta}_1 = (My_{2.T} - My_{1.T}) - (My_{2.C} - My_{1.C})$$

- أي نحسب الانتقال الزمني لكل من عينة المراقبة والمعالجة
- ثم نحسب الفرق بين الفرقين.

بتعويض متغيرتي التفسير بقيمهما في الدالة يمكن تمثيل نتائج التقدير وعادة ما يتم في جدول كالتالي:

	قبل	بعد	بعد ناقص قبل
عينة المراقبة	$\beta_0$	$\beta_0 + \delta_0$	$\delta_0$
عينة المعالجة	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
المعالجة - المراقبة	$\beta_1$	$\beta_1 + \delta_1$	$\delta_1$

في حالة إدراج عوامل أخرى في النموذج، لاحتساب الفروق الموجودة في العينة عبر الزمن، فإن قيمة المعامل  $\delta_1$  لا تحسب بالطريقة البسيطة (الفرق بين الفرقين) لكن تفسيره يبقى مشابهاً.

مثال<sup>1</sup> (مركز الردم). درس كاييل وماكلاين (Kiel et McClain, 1995) أثر وجود مركز ردم نفايات على سعر الشقق في نورث أندوفر North Andover، ماساشوستس، الولايات المتحدة. نشأت الإشاعة عن إنشاء المركز في 1978، وبدأ البناء في 1981، وبدأ المركز في العمل في 1985. البيانات المستخدمة هي لأسعار الشقق المباعة في 78 و82، وهما عيتان مستقلتان. الفرضية المختبرة هي أن الشقق المأجرة بقرب المركز (أقل من 3 أميال) أقل سعرا من الأبعد. سوف نستخدم السعر الحقيقي rprice عن طريق مؤشر السعر للعقار.

سوف نستخدم عدة دوال لكي نرى الاختلاف في التحليل وفائدة استخدام البيانات المجمعة، وتفسير المعاملات.

**الدالة الأولى:** استخدام بيانات 1981 فقط (بداية البناء)، وnearinc متغيرة وهمية ثنائية تأخذ 1 للشقق القريبة من المركز و0 للبعيدة.

$$\hat{rprice} = 101307.5 - 30688.27 \text{ nearinc}$$

$$(3092) \quad (5827.71) ;$$

$$N = 142; R^2 = 0.165.$$

• **فسر الثابت في الدالة أعلاه.** ج: الثابت هو متوسط سعر الشقق البعيدة عن المركز سنة 81، أي عند بداية البناء (هذا التفسير "الجيد" للثابت ممكن لأن النموذج لا يتضمن متغيرات مفسرة أخرى).

• **فسر معامل الانحدار الجزئي في الدالة.**

معامل الانحدار الجزئي: -30688.27 هو متوسط فرق السعر للشقق القريبة من المركز مقارنة بالبعيدة منه.

<sup>1</sup> Wooldridge J. M., p. 649.

• إختبر وجود فرق في السعر بين الشقق القريبة والبعيدة (سنة 81).

الجواب:  $t = 30688.27/5827.71 = 5.26$ ، وهي قيمة عالية، أكبر من القيمة الجدولية (لاختبار ثنائي القيمة الجدولية هي تقريبا 1.96)، وبالتالي يمكن أن نرفض فرضية تساوي متوسط سعر الشقق من الفئتين (ونستدل بالتالي على وجود فرق في السعر بين الشقق القريبة والبعيدة عن المركز).

• هل تدل النتيجة السابقة على أن خبر بناء المركز أدى إلى إنخفاض أسعار الشقق القريبة منه؟ وكيف يمكن التحقق من ذلك؟

الجواب: ليس بالضرورة، يمكن أن يكون المركز بني بالقرب من الشقق منخفضة السعر، وفي هذه الحالة ليس هو المسؤول عن تدني أسعارها لأنها كانت متدنية أصلا قبل بنائه. للتحقق من ذلك يتعين اختبار الفرق بين الأسعار قبل وجود إشاعة البناء، أي استخدام بيانات 78.

• الدالة الثانية: بيانات سنة 78 أعطت الدالة التالية:

$$\hat{rprice} = 82517.23 - 18824.37 \text{ nearinc}$$

$$(2653.79) \quad (4744.59)$$

$$N = 179; R^2 = 0.082.$$

• فسر الثابت في الدالة أعلاه. ج: الثابت هو متوسط سعر الشقق البعيدة عن المركز سنة 78، أي قبل بداية البناء (مرة أخرى هذا التفسير "الجيد" للثابت ممكن فقط لأن النموذج لا يتضمن متغيرات أخرى).

• هل كانت أسعار الشقق القريبة من المركز منخفضة مقارنة مع البعيدة عنه حتى قبل إنتشار إشاعة إنشائه؟ كيف يمكن اختبار ذلك؟

الجواب: نعم، حتى قبل إشاعة بناء المركز كانت أسعار الشقق المعنية منخفضة بفارق \$-18824.37\$ (وهو معامل الانحدار في الدالة أعلاه) لاختبار الفرق، نقسم المعامل على الخطأ المعياري، فنجد:  $t = -3.97 = -18824.37/4744.59$ ، وهي قيمة بعيدة عن  $-1.64$ ، إذن الفرق دال بمستوى معنوية 5% (اختبار أحادي وليس ثنائي لأننا نختبر الانخفاض وليس الاختلاف).

• هل ارتفعت أسعار الشقق البعيدة عن المركز أم انخفضت بين 78 و81؟

الجواب: إرتفعت من متوسط 82517.23 إلى 101307.5 أي بفارق 18790.27

• هل ارتفعت أسعار الشقق القريبة من المركز أم انخفضت بين 78 و81؟

الجواب: أسعار الشقق القريبة من المركز أيضا ارتفعت:

في 78 متوسط سعرها:  $(82517.23 - 18824.37 = 63692.86)$

في 81 متوسط سعرها:  $(101307.5 - 30688.27 = 70619)$ .

لكن مقدار ارتفاع الشقق القريبة  $(70619 - 63692.86 = 6926.14)$  هو أقل من الارتفاع الذي عرفته الشقق البعيدة عن مركز الريم  $(18790.27)$ .

• كيف يمكن من الدالتين أعلاه حساب فرق السعر العائد لبناء المركز معزولا عن فرق المنطقة وفرق الزمن؟

الجواب هو فرق الفروق، ويمكن حسابه بطريقتين:

ط1. متوسط الفرق بين ارتفاع الشقق القريبة وارتفاع الشقق البعيدة:

$$\hat{d} = 6926.14 - 18790.27 = -11863.9$$

ارتفاع الشقق القريبة اقل ب 11863.9 من ارتفاع الشقق البعيدة.

تسمى هذه القيمة **مقدر فرق الفروق** estimateur de la différence des différences

هذا يعني أن وجود المركز ربما أثر على أسعار الشقق القريبة منه بأن حد من ارتفاع أسعارها مقارنة مع الشقق البعيدة. المشكلة هي أننا لا نستطيع أن نعرف إن كان هذا التأثير دالا إحصائيا، لأن الطريقة لا تعطينا الخطأ المعياري للمقدر  $\hat{d}$ .

ط2: متوسط الفرق بين اختلاف السعر في 78 واختلافه في 81:

في 78 الاختلاف بين المنطقتين كان 18824.37 وفي 81 صار 30688.27، نحسب الفارق بين الفرقين،

$$\hat{d} = 18824.37 - 30688.27 = -11863.9$$

من الواضح أن الفارق في السعر بين المنطقتين زاد عما كان عليه قبل بناء المركز، أي أن مركز الردم كان له أثر على أسعار الشقق القريبة منه. المشكلة في الطريقة هي دائما عدم القدرة على الاختبار الإحصائي لهذا الأثر، وهنا تظهر فائدة نموذج البيانات المقطعية المجمعة.

• **كيف يمكن اختبار إن كان هذا الفرق (الفرق العائد للمركز) دالا إحصائيا؟**

الجواب. حصلنا على الفرق الذي يقيس الأثر الصافي لبناء المركز. بقي أن نختبر هل هو دال، نحتاج إلى إدراج متغيرة جديدة في الدالة هي التفاعل بين الزمن والمنطقة، ومن ثم نقسم معامل هذه المتغيرة على خطئه المعياري.

• **أكتب الدالة التي تعطي تأثير بناء المركز على أسعار الشقق بمعزل عن المنطقة والزمن، وفسر معاملاتها، واختبر تأثير بناء المركز.**

الدالة الثالثة: ندخل بيانات **مجمعة 78 و 81**، وندرج التفاعل بين الزمن والمنطقة كمتغيرة جديدة:

$$\hat{rprice} = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 (y81 * nearinc)$$

$\beta_0$ : متوسط سعر الشقق في 78 في المنطقة البعيدة عن المركز.

$\delta_0$ : الزيادة في أسعار الشقق البعيدة عن المركز خلال الفترة بين 78 و 81.

$\beta_0 + \delta_0$ : متوسط أسعار الشقق البعيدة عن المركز بعد البناء أي في 81.

$\beta_1$ : الفرق في أسعار الشقق بسبب المنطقة (القرب من المركز) بصرف النظر عن البدء في بناء المركز (independamment de l'existence de l'incinérateur).

$\delta_1$ : **التغير في أسعار الشقق بسبب المنطقة (وهو المطلوب)** عند افتراض أن أسعار الشقق القريبة والبعيد لم تتطور بشكل مختلف لأسباب أخرى (لأننا لم ندخل مفسرات أخرى). الدالة جاءت كالتالي:

$$\hat{rprice} = 82517.23 + 18790.29 y81 - 18824.37 nearinc - 11863.9 (y81 * nearinc)$$

$$(2727.91) \quad (4050.07) \quad (4874.32) \quad (7456.65)$$

$$N = 321, R^2 = 0.174,$$

لاحظ أولاً أن حجم العينة هو مجموع حجمي عينتي النموذجين السابقين  $321 = 179 + 142$

لاحظ أيضاً أن قيمة المقدر  $\delta_1 = -11863.9$  هي ذاتها قيمة  $\delta_d$ . الفائدة هنا هي أنه لدينا الخطأ المعياري لهذا المقدر، أي  $SE(\delta_1)$ ، وبالتالي يمكن استخدامها لاختبار تأثير بناء المركز على أسعار الشقق القريبة منه:

$$t = \delta_1 / SE(\delta_1) = -11863.9 / 7456.65 = -1.59$$

هذه القيمة لإحصائية الاختبار  $-1.59$  هي أقل من القيمة الجدولية لاختبار أحادي عند مستوى معنوية 5 بالمائة  $(-1.64)$ ، وبالتالي نقول بأن المعامل  $11863.9$  غير دال في اختبار وجود تأثير سلبي لوجود المركز على أسعار الشقق القريبة منه  $(pvalue = 0.057)$ . بعبارة أخرى، حسب نتيجة هذا الاختبار، لا يمكن الاستدلال على وجود تأثير سلبي لوجود المركز على أسعار الشقق القريبة منه (ولا نقول "تستدل على عدم وجود تأثير" ! ليس الأمر سيان !، ولذلك يمكن إجراء اختبارات أخرى، من خلال تعديل النموذج مثلاً بإدراج مفسرات أخرى).

• هل يمكن أن تتغير نتيجة الاختبار في حالة إدراج متغيرات أخرى؟ وهل إدراج متغيرات أخرى مهم؟

قام كايل وماكلين (1995) بإدخال خصائص متنوعة لسوق العقار (عدد الغرف، المساحة، ...)، هذه بالإضافة مبررة لسببين: أولاً لأن أنواع الشقق المباعة في 78 قد تكون مختلفة عن الأنواع المباعة في 81، وبالتالي يجب أخذ هذا في الحسبان. ثانياً، حتى إن لم تكن خصائص الشقق المباعة تغيرت (تأثيرها غير دال)، فإن إدخالها يمكن أن يقلص تباين الخطأ، مما يمكن أن يؤدي إلى تقليص الخطأ المعياري للمعامل  $\delta_1$ .

• علق على نتائج كايل وماكلين في النماذج الثلاث أدناه.

في النموذج الثاني تم إدخال عمر الشقق، وفي النموذج الثالث تم إدخال: العمر، عدد الغرف، عدد غرف الماء، المساحة المنزل، مساحة الأرض، المسافة عن حدود الولاية. المتغيرة التابعة هي السعر الحقيقي  $rprice$

المتغيرات	النموذج(1)	النموذج(2)	النموذج(3)
Constante	82517.23 (2727.91)	89116.54 (2406.05)	13807.67 (11166.59)
Y81	18790.29 (4050.07)	21321.04 (3443.63)	13928.48 (2789.75)
Nearinc	-18824.37 (4874.32)	-9397.97 (4812.22)	-3780.34 (4453.42)
Y81 * nearinc	-11863.90 (7456.65)	-21920.27 (6359.75)	-14177.93 (4987.27)
Autres facteurs	Non	age, age <sup>2</sup>	age, room, bath, surf, land, dist.
N	321	321	321
R <sup>2</sup>	0.174	0.414	0.660

Source : Cengage learning, 2013

كيف تغير  $R^2$  في الدالتين الموالتين؟

الجواب: كما هو متوقع، الدالتين الثانية والثالثة أعطتا نسبة تباين مفسر أكبر.

• ما هي نتيجة اختبار تأثير المركز في النموذجين الثاني والثالث؟

ج: في النموذج الثاني، على عكس النموذج الأول، جاء اختبار تأثير التداخل دالا:

$$t = -21920.27/6359.75 = -3.45$$

في النموذج الثالث كذلك:  $t = -14177.93/4987.27 = -2.84$

فسر التغير في المعامل **nearinc**!

ج: معامل **nearinc** (أثر المنطقة بصرف النظر عن إنشاء المركز) انخفض بشدة من النموذج الأول إلى

النموذجين الثاني والثالث، وهذا يعني أن العوامل الإضافية التي تم إدراجها لها تأثيرها.

• أي النماذج أفضل؟

النموذج الثالث أفضل لأنه يأخذ في الحسبان عوامل أكثر ولأن الخطأ المعياري لمعاملاته أقل (تقرير دقيق)، ما عدا بالنسبة للثابت، والذي ليس له أهمية هنا (ليس له بعد معنى اقتصادي مثل المعني الذي يكتسبه في النموذج الأول: متوسط سعر ... فهو مجرد ثابت طالما أن النموذج يتضمن متغيرات الخصائص مثل المساحة وغيرها).

• عودة إلى الدالة الثالثة: كيف يصبح التفسير لو حولنا المتغيرة التابعة **price** إلى اللغز بدلا من استخدام السعر الحقيقي **rprice**:

الجواب: تفسير  $\delta_1$  يصبح كما يلي:  $\delta_1(100)$  تعطي تقريبا نسبة الفرق في سعر الشقق العائد للمركز. باستخدام نفس المعطيات تأتي الدالة كما يلي:

$$\begin{aligned} \wedge \text{Log}(\text{price}) = & 11.29 + 0.457 y_{81} + 0.340 \text{nearinc} + 0.063 y_{81} * \text{nearinc} \\ & (0.31) \quad (0.045) \quad (0.055) \quad (0.083) \end{aligned}$$

$$N=321, R^2 = 0.409.$$

معامل حد التداخل يظهر أنه بسبب بناء المركز إنخفضت أسعار الشقق القريبة منه ب 6.3%، وهي غير دالة (أي نفس نتيجة الاختبار مع **rprice**. تم أيضا إدخال المتغيرات الأخرى في الدالة الأخيرة (الخصائص)، فأصبحت النسبة 13.2%.

ماذا لو استخدمنا **log(rprice)**؟

النتيجة هي ذاتها، الفرق فقط في الثابت: استخدام لغ السعر الإسمي **log(price)** أو لغ السعر الحقيقي **log(rprice)** لا يغير إلا الثابت فقط.

مثال 2. (مدة النقاهاة<sup>1</sup>). درس ماير، فيسكوسي، ودورين (Meyer, Viscusi, et Durbin ; 1995) المدة بالأسبوع للنقاهاة التي يحصل عليها العامل (تعويض) بعد حادث عمل (durat). رفعت ولاية كانتوكي في 1980 سقف الدخل الأسبوعي الذي يغطيه نظام تعويضات حوادث العمل للموظفين ذوي الدخل العالي، علما أن الإجراء لم يمس ذوي الدخل المنخفض. للقيام بالدراسة تم تكوين عينة مراقبة (ممن لا يمسه الإجراء)، وعينة معالجة (ممن يعينهم الإجراء أي من ذوي الدخل العالي). من خلال استطلاع عينة عشوائية من الموظفين قبل وبعد الإجراء، أراد

<sup>1</sup> Wooldridge J. M., p. 654.



الباحثون اختبار ما إذا كان المستفيدون منه استخدموا مدة نقاهة أطول بسبب الإجراء. قام الباحثون أولاً بتقدير نموذج فرق فروق باستخدام لغزتم المدة  $\log(\text{durat})$  كمتغيرة تابعة. المتغيرة  $\text{aftchnge}$  هي متغيرة ثنائية تعبر عن الفترة قبل أم بعد السياسة، والمتغيرة  $\text{highearn}$  تعبر عن فئة الدخل (1 = دخل عالي، 0 = لا). النتائج كانت كالتالي:

$$\hat{\log(\text{durat})} = 1.126 + 0.0077 \text{aftchnge} + 0.256 \text{highearn} + 0.191 \text{aftchnge} * \text{highearn}$$

$$\text{SE} = (0.031) \quad (0.0447) \quad (0.047) \quad (0.069)$$

$N = 5626; R^2 = 0.021.$

1. كم هي نسبة التغير في مدة النقاهة للفئة المستفيدة نتيجة للإجراء؟ هل التغير دال إحصائياً؟
2. كم هي نسبة التغير في مدة النقاهة لغير المعنيين بالإجراء؟ علق على قيمتها؟
3. كيف يفسر معامل  $\text{highearn}$ ؟
4. كم هي نسبة التباين المفسر بالنموذج؟ علق عليها.
5. علق على حجم العينة.

الجواب:

- $\delta^1 = 0.191$  تمثل نسبة التغير في مدة النقاهة لدى الفئة المستفيدة وهي 19% تقريباً، وهي نسبة معتبرة. بقسمة هذه القيمة على الخطأ المعياري نحصل على:  $(t = 0.191/0.069 = 2.77)$ ، الزيادة إذن دالة إحصائياً. (النسبة الحقيقية هي  $1 - \exp(0.191) = 0.21$  أي 21 بالمائة).
- معامل  $\text{aftchnge}$  يمثل أثر السياسة على غير ذوي الدخل العالي أي الذين لا يعينهم الإجراء (عينة المراقبة)؛ نسبته 0.77%، وهي نسبة شبه معدومة، وذلك طبيعي لأن الإجراء لم يمسه.
- معامل  $\text{highearn}$  يمثل الفرق في مدة الاجازة عند ذوي الدخل العالي، ويعني أن ذوي الدخل العالي يأخذون إجازات نقاهة أطول حتى قبل الإجراء المتخذ لفائدتهم ( $\text{aftchnge} = 0$ )، حوالي 29.2% في المتوسط:  $\exp(0.256) - 1 = 0.2917$
- نسبة التباين المفسر بالنموذج ضعيفة، 2.1%، وهذا طبيعي لأن العوامل التي تؤثر في مدة النقاهة عديدة، أهمها طبيعة وشدة الحادث، ... لحسن الحظ، كبر حجم العينة سمح بالحصول على نتائج دالة، رغم ضعف نسبة التباين المفسر. لاحظ أن الدراسة سمحت باستكشاف تأثير السياسة على التابع (مدة العطل المرضية) رغم أن معظم تباين هذه الأخيرة غير محتسب بالنموذج.
- للإشارة، أخذ الباحثون في نموذج آخر عدة عوامل أخرى، مثل الجنس، والعمر، والحالة العائلية، والقطاع، ونوع الجروح. إدراج هذه العوامل لم يكن له إلا تأثير خفيف على المعامل  $\delta_1$ .

رأينا في هذا الفصل أهمية البيانات المجمعة وكيف تستخدم لإظهار تطور المتغيرة التابعة في الزمن من خلال إدراج متغيرات وهمية ثنائية، وكيفية قياس تطور تأثير العوامل الأخرى من خلال تفاعل العوامل مع الزمن.

## فصل 2. نموذج الأثر الفردي وتقديره بطريقة POLS وطريقة FD

نموذج الأثر الفردي - طريقة POLS - طريقة الفروق الأولى لحالة الفترتين - طريقة الفروق الأولى لحالة أكثر من فترتين

بيانات البانل هي بيانات لعينة من المفردات (مؤسسات، مدن، بلدان، أفراد...) أخذت على فترتين أو أكثر، أي أن الباحث يرجع لنفس المفردات لأخذ القياسات أو البيانات، على خلاف البيانات المقطعية المجمعَة أين تكون العينة في كل فترة هي عينة جديدة أي مفردات جديدة. هناك أكثر من طريقة لمعالجة هذا النوع من البيانات. طرق تقدير نماذج البانل تختلف بحسب الاعتماد على التباين الداخلي أم التباين البيئي في البيانات أم كلاهما (كما في POLS). خصائص المقدرات تختلف بحسب أي النموذج المناسب. نفضل المقدرات المتقاربة والفعالة (consistent and efficient)، ونبحث أولاً عن المقدرات المتقاربة، أي حين تكون العينات الأكبر تعطي تقديراً أصدق وأكثر دقة. التقارب رياضياً هو عندما ينطبق أو ينكمش (collapse) التوزيع الاحتمالي للمقدر على المعلمة عندما تكون  $n$  ضخمة  $\text{plim } \hat{\beta}_n = \beta$ . يستدل على التقارب باستخدام نظرية الأعداد الكبيرة. أما الفعالية فهي متعلقة بتباين المقدر. في هذا الفصل نتطرق لطريقة الفروق الأولى من خلال أمثلة عن دراسات حقيقية. نتطرق أيضاً لحالة تعدد الفترات. الأمثلة والشروح مستقاة أساساً من المرجع المذكور: وولدرج Wooldridge وبعض المراجع الأخرى.

### 1. نموذج الأثر الفردي وتقديره بطريقة POLS

تنظيم البيانات  
نموذج الأثر الفردي  
التقدير بطريقة POLS

#### 1-1. مبرر بيانات البانل

من خلال المثال التالي نرى أهمية بيانات البانل وطريقة معالجتها.

**مثال الجريمة والبطالة:** لدى باحث بيانات معدل الجريمة  $crmrte$  (كم جريمة لكل 1000 ساكن) ومعدل البطالة  $unem$  لـ 46 مدينة أمريكية في 1982 و1987. باستخدام بيانات 1987 وحدها (بيانات مقطعية) أتت المعادلة كما يلي:

$$\hat{Crmrte} = 128.38 - 4.16unem$$

(20.76)    (3.42)

$$N = 46, R^2 = 0.033.$$

لاحظ أن معامل الانحدار الجزئي سالب! كأن العلاقة إن وجدت فهي عكسية: المدن التي تزيد فيها البطالة تقل فيها الجريمة والعكس! في الوقت نفسه الخطأ المعياري كبير (يكاد يساوي قيمة المعامل نفسه، مما يعني أن التقدير غير دقيق واختبار التأثير غير دال) وبالتالي فإن العلاقة غير مؤكدة. لم تثبت البيانات إذن وجود علاقة. نسبة التباين المفسر أيضا ضئيلة جدا! الواقع أن هناك العديد من المتغيرات التي تؤثر على معدل الجريمة لم تدرج، ولذلك نشك أن المقدرات منحازة. إدراج مفسرات أخرى يمكن أن يكون أحد الحلول لمشكلة المفسرات المهملة وهو يسمح بتحسين نسبة التباين المفسر. من هذه المفسرات التي يمكن إدراجها: نسبة الشباب في المدينة، نسبة الذكور، صرامة القوانين، جهود الأمن وفرض القوانين، مستوى الدخل في المدينة، المستوى التعليمي في المدينة... غير أن إدراج هذه المتغيرات يمكن أن يكون صعبا، لصعوبة قياس هذه المتغيرات أو لعدم توفر البيانات ... فهل من بديل؟

الخلاصة أن استخدام بيانات مقطعية (وليس بيانات البائل) هناك متغيرات أهملت في النموذج وهذا من شأنه أن يجعل المقدرات (معاملات الدالة) متحيزة (مقدر في المثال سالب، بينما نعلم أن التأثير طردي). فما هو الحل؟ الحل هو أن نأخذ سوابق المدينة في الجريمة  $crmrte$  في النموذج، مما يكشف توجهات تاريخية تفسر لماذا معدل الجريمة أعلى في مدن معينة مقارنة بالأخرى. هذا أحد استخدامات بيانات البائل: حل مشكلة المتغيرات المهملة التي من شأنها أن تسبب التحيز في المقدرات.

بفرض أن لدينا 50 مدينة، مشاهدة على مرحلتين. هذا يعني أنه سيكون لدينا 100 سطر. الطريقة الأولى والشائعة لتنظيم البيانات هي تخصيص سطرين لكل مدينة، السطرين الأولين للمدينة الأولى في العينة: الأول للمشاهدة الأولى والثاني للمشاهدة الثانية، والسطرين المواليين للمدينة الثانية، وهكذا. بهذه الطريقة يسهل حساب الفروق بين السنتين، كما يسهل استخدام طريقة البيانات المجمعلة للمقارنة بين الطريقتين: التقدير بالبيانات المجمعلة والفروق الأولى. في حالة وجود أكثر من فترتين، يمكن استخدام ذات الطريقة: المشاهدات الخاصة بكل مفردة (مدينة) مرتبة تصاعديا بجوار بعضها.

الطريقة الثانية هي تخصيص سطر واحد لكل مفردة (مدينة) وهذا يعني إدخال كل متغيرة مرتين، مرة لكل فترة. في مثال النوم والعمل سيكون لدينا لكل عامل المتغيرات التالية:  $slpnap75, slpnap81, totwrk75, totwrk81, \dots$  بحيث تحسب الفروق أفقيا لا عموديا. هذه الطريقة لا تسمح باستخدام طريقة التقدير بالبيانات المجمعلة، كما أنها لا تصلح عندما تكون الفترات أكثر من 2.

## 2-1. نموذج الأثر الفردي *Individual-specific effects model*

هناك عدة نماذج في بيانات البائل وعدة طرق لتقدير معالمها. فيما يلي نتطرق لنموذج الأثر الفردي وطريقة الفروق الأولى<sup>1</sup> في تقدير معالم هذا النموذج، مع تطبيق على تقييم سياسة عمومية. في نموذج الأثر الفردي (أو بالأحرى الأثار الفردية أو الأثار غير المشاهدة)، نفترض أن هناك فروقا فردية غير مشاهدة (unobserved heterogeneity) ممثلة بـ  $\alpha_i$ . إذا أخذنا مثلا تأثر الأجر بالخبرة. سنجد أن الأجر يتأثر أيضا بمتغيرات أخرى منها معامل الذكاء، وعدد سنوات الدراسة، وهذه تختلف من فرد لآخر لكنها تقريبا ثابتة في الزمن. يطرح هنا التساؤل:

<sup>1</sup> FD : First Differences (Différences premières).

هل  $\alpha_i$  مرتبطة بالمفسرات أم لا، إذا كانت مرتبطة نذهب إلى طريقة الأثر الثابت، إذا لم تكن مرتبطة نستخدم طريقة الأثر العشوائي. طريقة الأثر الثابت FE تسمح بارتباط الفروقات الفردية بالمفسرات، ويضع لكل مقدر ثابتا خاصا  $\alpha_i$ ، ولكن نفس الميل  $\beta$ . يمكن أيضا إدراج متغيرات وهمية للزمن (time dummies).

المتغيرات المفسرة يمكن تقسيمها إلى صنفين: متغيرات ثابتة في الزمن، وأخرى متغيرة. كيف نأخذ بالاعتبار في النموذج هذا التقسيم؟

**نموذج الآثار الفردية أو غير المشاهدة (Individual-specific Effects Model or UEM: Unobserved Effects Model)** يأتي كما يلي:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it} + \alpha_i + u_{it}, \quad t = 1, 2.$$

- حيث  $i$  تمثل المفردة، و  $t$  تمثل الزمن.
- متغيرة ثنائية، تأخذ 1 لما  $t = 2$ ، و 0 لما  $t = 1$ .
- $x_{it}$  هي المتغيرة المفسرة (في المثال هي معدل البطالة)<sup>1</sup>
- الثابت إذن هنا، يأخذ  $\beta_0$  في الفترة الأولى، و  $(\beta_0 + \delta_0)$  في الفترة الثانية. بهذا نكون قد احتسبنا التوجه التاريخي للمتغيرة التابعة. في المثال لابد أن معدل الجريمة يتغير مع الزمن زيادة أو نقصانا، في عموم البلاد، فالغالب أن هناك توجهها تصاعديا للجريمة يجب أخذه في الاعتبار.
- $\alpha_i$  هو الأثر الثابت في الزمن (لا وجود عليه للمؤشر  $t$ )، فهو يتحرك على بعد واحد هو البعد الفردي. هذا المعامل يمثل الفروقات الفردية غير المشاهدة unobserved heterogeneity or unobserved component or (latent variable) بين المدن التي لا تتغير في الزمن، والتي تؤثر على الجريمة: مثلا متغيرات جغرافية أو ديمغرافية ثابتة زمنيا، مثل التوزيع العرقي في المدينة، أو المستوى الدراسي العام، ... يسمى  $\alpha_i$  أيضا الأثر غير المشاهد.  $\alpha_i$  يمثل إذن متغيرة تتغير على المستوى الفردي ولكنها ثابتة في الزمن.
- الخطأ  $u_{it}$  نسميه الخطأ البسيط أو الذاتي idiosyncratic error or idiosyncratic disturbances، وهو خطأ متغير في الزمن، فهو يمثل عوامل الجريمة المتغيرة زمنيا.

**لاحظ.**

النقاش حول هل يتم اعتبار  $a_i$  أثرا ثابتا أم عشوائيا هو في الأصل هل هذا الحد هو فعلا متغيرة عشوائية أم معلمة يتعين تقديرها. لكن ما يهم حسب وولدرج هو هل هذا الحد مرتبط بالمفسرات أم لا. طريقة الأثر العشوائي RE تفترض عدم وجود ارتباط  $Cov(x_{it}, a_i) = 0$  وفي الواقع فإن صيغة أشد للفرضية stronger hyp، بمعنى الاستقلال، هي المطلوبة لتبرير الاستدلال تبريرا كاملا، وهي  $E(a_i/x_{it}) = E(a_i)$ . في أدبيات الاقتصاد القياسي، مصطلح الأثر الثابت لا يعني بالضرورة أن  $a_i$  يتم اعتبارها متغيرة غير عشوائية، وإنما أن الباحث يسمح بوجود ارتباط ما اعتباطي بين الأثر غير المشاهد والمتغيرات المفسرة. لذلك فإن اصطلاح "الأثر الفردي الثابت" أو "الأثر الثابت الصارم" (firm fixed effect) يعني أن الأثر غير المشاهد يسمح له بالارتباط بالمفسرات<sup>2</sup>. لذلك فالصحيح حسب وولدرج هو الإشارة إلى  $a_i$  ليس كأثر ثابت أو عشوائي وإنما

<sup>1</sup> يمكن أن نتصور عدة عوامل أو أيضا عدة متغيرات وهمية تعبر عن فترات زمنية متعددة، يمكن أيضا أن نتصور وجود عامل في النموذج يتغير في الزمن فقط  $Tt$  يمس المتغيرة التابعة ليغيرها باتجاه معين مع الزمن، مثلا: أسعار الشقق تزداد مع الزمن بسبب ازدياد ثروة ودخل الفرد في البلد.

<sup>2</sup> Wooldridge, Econometric analysis of cross section and panel data, MIT Press, USA, p. 252.

كأثر غير مشاهد أو تباين غير مشاهد ... unobserved effect, unobserved heterogeneity, ... ومع ذلك فإن مصطلح FE و RE شائعين في الأدبيات لدرجة أنه لا يمكن تغيير ذلك الآن حسب الكاتب.

**كيف يمكن تقدير  $\beta_1$  ؟ هناك ثلاث أنواع رئيسية من النماذج: The pooled model, the fixed effects model, the random effects model**

### 1-3. طريقة POLS

من بين الطرق أو النماذج لمعالجة بيانات البانل ما يسمى المربعات الصغرى على مجمل البيانات Pooled OLS أو Pooled، أو أيضا <sup>1</sup>POLS. وهي طريقة نادرة الاستخدام لعيوبها، وهي غالبا ما تذكر لتبيان أفضلية النماذج الأخرى. في طريقة Pooled يحتسب الزمن بمتغيرة ثنائية، فيأتي النموذج:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it} + v_{it}, \quad t = 1, 2.$$

حيث  $v_{it} = a_i + u_{it}$  ويسمى الخطأ المركب l'erreur composée

العيب الأساسي لهذه الطريقة هو أنه لتجميع البيانات للحصول على مقدر متقارب ل  $\beta_1$ ، يجب افتراض أن الأثر الثابت  $a_i$  ليس مرتبطا ب  $x_{it}$  سواء من أجل  $t = 1$  أو  $t = 2$ . حتى إذا كانت  $u_{it}$  غير مرتبطة ب  $x_{it}$  فإنه إذا كانت  $a_i$  مرتبطة ب  $x_{it}$  فإن طريقة المربعات الصغرى على بيانات البانل تعطي مقدرات متحيزة و(الأهم من ذلك) غير متقاربة (inconsistent). يسمى هذا التحيز تحيز الفروقات الفردية heterogeneity bias وهو عائد لإهمال متغيرة ثابتة في الزمن.

شرط عدم الارتباط هذا ضروري أيضا لتقدير المعالم الأخرى للنموذج (وهو شرط سواء استخدمنا بيانات مقطعية أو مجمعة). لكن في أغلب الأحيان هذا الشرط غير محقق؛ في مثال الجريمة والبطالة في المدن فإن متغيرات ثابتة في الزمن مثل المعدل العام للعمر في المدينة أو المستوى التعليمي العام يحتمل جدا أن تكون لها علاقة مع البطالة أيضا أي المتغيرة المفسرة. وجود الارتباط بين حد الخطأ والمتغيرات المفسرة يعبر عنه ب:

$$\text{Cov}(v_{it}, x_{it}) \neq 0$$

من جهة أخرى، فإن الفائدة الأساسية من بيانات بانل هو وجود ارتباط بين المتغيرات المفسرة والأثر الثابت  $a_i$ .

إدراج متغيرة ثنائية للزمن لا يحل المشكلة: للتوضيح نعود لبيانات المثال ونجمع هذه المرة بيانات الفترتين. سيكون لدينا 46 مضمروبا في 2، أي 92 مشاهدة.

$$\text{Crmrte}^{\wedge} = 93.42 - 7.94d87dT2 + 0.427unem$$

$$(12.74) \quad (7.98) \quad (1.188)$$

$$N = 92, \quad R^2 = 0.012.$$

<sup>1</sup> يسمى هذا النموذج في (pooled) R. والدالة المستخدمة للقيام بالتحليل هي (plm). يمكن أيضا استخدام متوسط مقدرات الفترة لكل مفردة، ومن ثم إجراء تحليل الانحدار. تسمى هذه الطريقة في R (between).

<sup>2</sup> يفترض أيضا طبعا تحقق الاستقلال من أجل أي قيمة ل  $a_i$ ، ونكتب  $\text{Cov}(v_{it}, x_{it}) = 0$  For any  $i$  or  $t$ ، ولتكون مقدرات المتوسطات المتحركة غير متحيزة نحتاج إلى شرط الاستقلال ليس فقط من أجل نفس التاريخ، ونكتب:  $\text{Cov}(v_{it}, x_{is}) = 0$

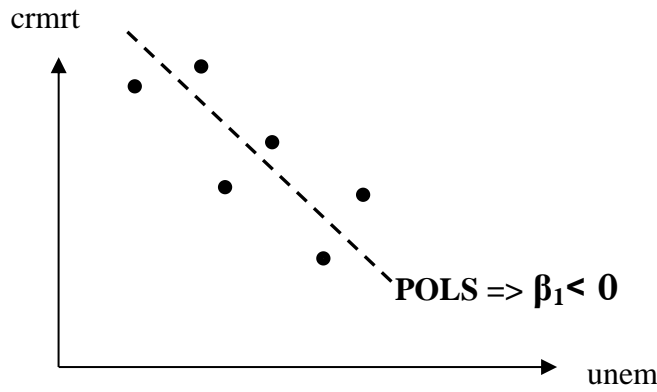
معامل **unem** غير دال. استخدام POLS إذن لم تحل المشكلة وهذا متوقع فهي لا تحل مشكلة المتغيرات المهملة (في الواقع الخطأ المعياري المقدر هنا غير صحيح بسبب الارتباط الذاتي ... serial correlation).

لشرح تأثير الفروقات الفردية (متغيرات ثابتة في الزمن) على معامل المفسر، نأخذ المثال السابق مع تبسيط أكبر بحيث نستغني عن المتغيرة الثنائية للزمن.

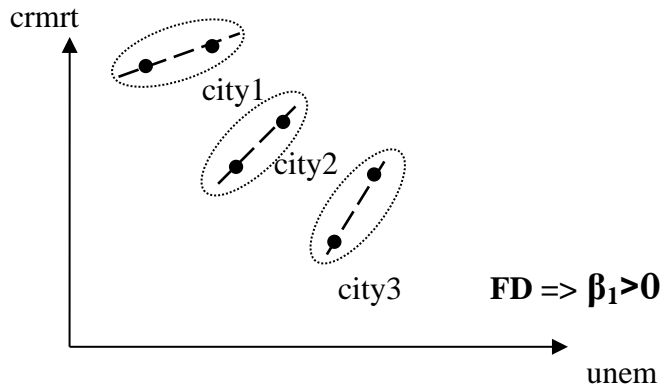
**مثال البطالة والجريمة.** ولنفترض أن محددات الجريمة في مدينة  $i$  في لحظة  $t$  هي البطالة فيها وخصوصية للمدينة ثابتة في الزمن  $\alpha_i$  ومجموعة من العوامل الأخرى  $u_{it}$ . لنفترض أننا أخذنا بيانات سنتين واستخدمنا طريقة المربعات الصغرى على مجمل البيانات بدون إدراج متغيرة للزمن، ولنفترض للتبسيط أن لدينا فقط 3 مدن وبالتالي 6 مشاهدات. النموذج هو كالتالي:

$$Crmrt_{it} = \beta_0 + \beta_1 unem_{it} + \alpha_i + u_{it}, \quad t = 1, 2.$$

يمكن أن يأتي الرسم كالتالي:



لاحظ أن POLS أعطت هنا معامل سالب!! لأنها تعتبر المشاهدة وليس المدينة. بينما في الواقع يتعلق الأمر ب 3 مدن، وفي كل مدينة لدينا مشاهدتين، واحدة للسنة الأولى والثانية للسنة الثانية. ويمكن إذا نظرنا إلى البيانات من هذه الزاوية أن نجد أن العلاقة في الواقع طردية في كل من المدن الثلاث (أنظر الرسم أدناه).



كخلاصة: طريقة Pooled، لها عيب أساسي هو **تحيز الفروق الفردية** أو ما يسمى أيضا بالفروق الفردية غير المشاهدة (unobserved heterogeneity)، والذي ينجم عن عدم استقلال حد الخطأ في النموذج عن العامل (أو العوامل). هذه الطريقة قل ما نجدها مستخدمة، فهي تتجاهل أن البيانات هي بيانات بانل.

**لاحظ:** طريقة المقدر البيني. هناك أيضا طريقة أخرى قليلة الاستخدام وهي OLS على المتوسطات في الزمن، وتسمى طريقة المقدر البيني Between estimator حيث يتم تحليل الانحدار على متوسطات التابع ومتوسطات المتغيرات المفسرة،

بحيث نأخذ لكل مفردة متوسط التابع في الزمن ومتوسط المفسر أو المفسرات في الزمن. حجم العينة في هذه الحالة يصبح  $n$ . عيب هذه الطريقة هو أنها تهمل التباين في الزمن للمتغيرات وتقلص حجم العينة. هذه الطريقة قليلة الاستخدام لأن مقدرات نموذج RE و POLS أكثر فعالية منها. من البدائل الحلول لتجاوز مشكلة الفروقات الفردية غير المشاهدة طريقة الفروق الأولى (First Differences).

## 2. دالة الفروق الأولى

تعريف بالطريقة  
شروط دالة FD  
عوائق الطريقة  
حالة تعدد المتغيرات المفسرة  
النماذج ذات التأخرات  
تطبيق على تقييم السياسات العمومية  
دالة الفروق الأولى لأكثر من فترتين

### 1-2. تعريف بالطريقة

دالة الفروق الأولى هي طريقة سهلة لاحتساب المتغيرات المهملة في بيانات البانل، وتحل مشكلة ارتباط الفروق الفردية الثابتة أو الأثر الثابت بالمفسرات، علما أنه في أغلب الأحيان الفائدة الأساسية من بيانات بانل هو وجود ارتباط بين المتغيرات المفسرة والأثر الثابت  $a_i$ . في مثال الجريمة وتأثرها بالبطالة، نريد احتساب أن هناك علاقة بين معدل البطالة (المتغيرة المستقلة) والفروقات الكامنة في  $a_i$  مثل التوزيع الديمغرافي في المدينة (نسبة الشباب، نسبة النساء...)، المستوى العام للدخل في المدينة، ...

دالة الفروق الأولى لا تفترض أن الأثر الثابت  $a_i$  ليس مرتبطا ب  $x_{it}$ : بما أن  $a_i$  لا يتغير مع الزمن، يكفي أن نحذفه من النموذج بطرح سنة من أخرى:

$$y_{i2} = \beta_0 + \delta_0 (1) + \beta_1 x_{i2} + \alpha_i + u_{i2},$$

$$y_{i1} = \beta_0 + \delta_0 (0) + \beta_1 x_{i1} + \alpha_i + u_{i1}$$

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + (u_{i2} - u_{i1}),$$

أو أيضا:

$$\Delta y_{it} = \delta_0 + \beta_1 \Delta x_{it} + \Delta u_{it}$$

تسمى هذه الدالة دالة الفروق الأولى.

$\delta_0$  هو تغير الثابت بين الفترتين؛ إذا لم يكن معدوما، يعطي  $\delta_0$  انتقال  $y$  في الزمن عند ثبات  $x$  في الزمن، بعبارة أخرى هو يعطي التوجه الزمني العام (للمتغيرة التابعة في المجتمع). إشارة الثابت تبين إن كان التوجه هو الزيادة أم النقص.

مقدر  $\beta_1$  المحصل عليه بطريقة المربعات الصغرى يدعى مقدر الفروق الأولى. تفسيره يبقى التفسير الأصلي لمعامل الانحدار الجزئي: مقدار التغير في  $y$  عند تغير  $x$  بوحدة واحدة<sup>1</sup>.

**مثال الجريمة ثانية:** باستخدام دالة الفروق الأولى أعلاه في مثال الجريمة نحصل على:

$$\Delta Crmrte^A = 15.40 + 2.22 \Delta unem$$

$$(4.7) \quad (0.88)$$

$$N = 46, R^2 = 0.127.$$

المعامل  $\beta_1$  يختلف معنويا عن الصفر (بقسمته على خطئه المعياري نجد قيمته أكبر من 2). العلاقة إذن دالة وطرديّة بين معدل البطالة ومعدل الجريمة في المدينة. الثابت أيضا له هنا أهميته: عندما تكون  $\Delta unem$  تساوي 0، أي عند ثبات معدل البطالة، فإن معدل الجريمة يزيد بـ 15.4 جريمة في 1000 ساكن. هذا يشير إلى وجود توجه صاعد للجريمة في البلاد بين 1982 و1987.

**الخلاصة:** رأينا هنا أهمية دالة الفروق الأولى<sup>2</sup>: بدلا من استخدام بيانات مقطعية يصعب معها تفسير أثر العامل عند تثبيت العوامل الأخرى، بسبب إهمال هذه الأخيرة، فإن النموذج أعلاه يعطينا بوضوح كيف يؤثر التغير في العوامل على التغير في  $y$ ، وكيف تتغير هذه الأخيرة في الزمن عند ثبات هذه العوامل. أثر  $x_{it}$  على  $y_{it}$  تم قياسه من خلال تثبيت  $a_i$  في الزمن.

## 2-2. شروط دالة الفروق الأولى

كل ما يميز هذه الدالة هو أن المتغيرات متفاضلة (differentiated)، ويمكن استخدام طريقة المربعات الصغرى عليها بشرط أن تكون الفرضيات الأساسية محققة، وهي ثلاث:

1- أهمهما أن تكون  $\Delta x_i$  و  $\Delta u_i$  غير مرتبطة. تكون هذه الفرضية محققة إذا كانت  $u_{it}$  لكل فترة غير مرتبطة بالمتغيرة المفسرة<sup>3</sup>. لكن لا يشترط أن تكون  $x_{it}$  مستقلة عن المتغيرات غير المشاهدة الثابتة مع الزمن.

في مثال الجريمة، القول بأن  $\Delta unem_i$  و  $\Delta u_i$  غير مرتبطان يمكن أن يكون صحيحا، كما يمكن أيضا أن يكون غير صحيح. مثلا، جهود الشرطة في فرض القانون (موجودة ضمن  $u_{it}$ ) قد تكون تزيد أكثر في المدن التي ينخفض فيها معدل البطالة، هذا الارتباط بين  $\Delta unem_i$  و  $\Delta u_i$  يخلق تحيزا في مقدرات MCO. هذا المشكل يمكن أن يعالج جزئيا بإدراج متغيرات مفسرة أخرى (متغيرة في الزمن) في النموذج.

2- الفرضية الثانية التي يجب أن تتحقق هي أن  $\Delta x_i$  تتباين كفاية من مشاهدة لأخرى. هذا الشرط لا يتحقق إذا كانت المفسرة لا تتغير في الزمن (متغيرة جغرافية مثلا...) أو إذا كانت تتغير بين الفترتين بنفس القدر لجميع المشاهدات. في مثال الجريمة نعلم أن معدل البطالة يتغير في الزمن بالنسبة لكل المدن تقريبا.

<sup>1</sup> (مبدئيا)، في حالة كون  $Cov(\Delta u_i, \Delta x_i) = 0$ ، وإذا لم يكن هناك عدم تجانس no heteroscedasticity فإن طريقة المربعات الصغرى تعطي

مقدرات متقاربة، وإذا حدث أيضا أن لم يكن هناك ارتباط ذاتي في حد الخطأ  $\Delta u_{it}$  عندها تكون المقدرات فعالة أيضا، في حالة  $T = 2$  فإن

نموذج FD يكافئ حسابيا نموذج FE

<sup>2</sup> يسمى هذا النموذج في R (fd).

<sup>3</sup> وهي شكل آخر لفرضية الاستقلال الصريح (exogénéité stricte)



3- الفرضية الثالثة هي **فرضية التجانس** (homoscedasticity)، وهي فرضية معقولة في كثير من الحالات، وهناك طرق لاختبارها والتصحيح في حالة **عدم التجانس** (heteroscedasticity).

أحيانا يمكن افتراض أن كل شروط النموذج الخطي التقليدي محققة، مقدرات المربعات الصغرى تكون غير متحيزة، وفي هذه الحالة فإن الاستدلال الإحصائي يكون صحيحا.

## 2-3. عوائق طريقة الفروق الأولى

طريقة الفروق الأولى إذن قوية في احتساب العوامل غير المشاهدة، لكن لها حدودها:

1- عيب أو عائق آخر في طريقة الفروق الأولى يكمن في أنها يمكن أن تقلل تباين المتغيرة المفسرة. حتى عندما تكون المفسرة متباينة في كل فترة، يمكن أن تكون  $\Delta x_i$  قليلة التغير، مما يجعل التقدير بطريقة المربعات الصغرى غير دقيق كفاية؛ فكلما كان تباين  $\Delta x_i$  ضعيفا كلما كان تباين المقدر ل  $\beta_1$  عاليا. أحيانا يمكن مواجهة هذا العائق باتخاذ عينة كبيرة الحجم أو باتخاذ مسافة زمنية كافية بين الفترتين، مثلا عدة سنوات تفصل بين المشاهدات بدل سنة واحدة.

2- يصعب في كثير من الأحيان الحصول على بيانات البانل على فترة طويلة، فالمؤسسات تفضل أو تدمج، أو تنتقل، والأفراد يموتون أو يرحلون...، (لكن تصلح أكثر مع الدول).

**مثال 2 مردودية الدراسة.** عند تقدير مردودية سنوات الدراسة على الأجر، من خلال بيانات سنتين لنفس العينة. أكتب نموذج الأثر الثابت مع تحويل التابع إلى اللوغاريتم. يكون النموذج كالتالي:

$$\log(\text{wage}_{it}) = \beta_0 + \delta_0 d_{2t} + \beta_1 \text{educ}_{it} + \alpha_i + u_{it}, t = 1, 2.$$

إشرح مدلول  $a_i$  وأذكر بعض العوامل التي قد تدخل في تكوينه، وهل يمكن القول أنها غير مرتبطة بالمفسر؟ المعامل  $a_i$  يمثل الفروقات الفردية التي تؤثر على الأجر وهي ثابتة في الزمن، مثل الذكاء والذاكرة والمهارات الأخرى، وهي في الغالب مرتبطة بالمفسرة عدد سنوات الدراسة  $\text{educ}$ ، لأن الأفراد الأكثر نكاء مثلا قد يكونون أكثر تعلما.

هل يمكن القول أن معادلة الفروق الأولى صالحة هنا؟ أكتبها.

بما أن الفروق الفردية  $a_i$  (الذكاء، الذاكرة...) قد لا تتغير في الزمن، فإن طريقة بيانات البانل تبدو مناسبة لتقييم مردودية الدراسة. معادلة الفروق الأولى تأتي كما يلي:

$$\Delta \log(\text{wage}_i) = \delta_0 + \beta_1 \Delta \text{educ}_i + \Delta u_i$$

ما فائدة المعامل  $\delta_0$  هنا؟

ندرج التغير في الثابت لاحتمال التغير في مردودية الدراسة عبر الزمن.

هل نتوقع أن تعطي طريقة المربعات الصغرى هنا مقدرات دقيقة؟

المشكلة هي أننا معنيون بالراشدين، وهؤلاء غالبا لا تتغير عدد سنوات دراستهم من سنة لأخرى إلا لقله منهم وبقدر قليل، وفي هذه الحالة فإنه يصعب تقدير  $\beta_1$  بدقة، إلا إذا كان لدينا عينة كبيرة جدا.

## 4-2. حالة تعدد المتغيرات المفسرة

رأينا أنه في حالة متغيرة وحيدة مفسرة يكون النموذج كما يلي:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it} + \alpha_i + u_{it}, \quad t = 1, 2.$$

ودالة الفروق الأولى في حالة متغيرة وحيدة تكتب كما يلي:

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

إضافة متغيرات إلى النموذج سهل، ويكتب النموذج كما يلي:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + \alpha_i + u_{it}, \quad t = 1, 2.$$

المؤشر الثالث هو لرقم المتغيرة.

دالة الفروق الأولى تصبح كما يلي:

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_{i1} + \beta_2 \Delta x_{i2} + \dots + \Delta u_i$$

مثال 3: الموازنة بين الراحة والعمل.

درس بايدل و همرماش (1990) Biddle and Hamermesh العلاقة بين مدة النوم بالساعة  $\text{snap}$  والعمل  $\text{totwrk}$  بالساعة أيضا. بيانات البانل لسنتي 1975 و 1981 تخص 239 فرد. المفسرات الأخرى هي عدد سنوات الدراسة  $\text{educ}$ ، ومتغيرات ثنائية: متزوج أم لا  $\text{mar}$ ، وجود رضيع  $\text{yngkid}$  وبصحة جيدة  $\text{gdhlth}$ . يمكن استخدام نموذج الأثر الثابت التالي:

$$\text{snap}_{it} = \beta_0 + \delta_0 d_{1t} + \beta_1 \text{totwrk}_{it} + \beta_2 \text{educ}_{it} + \beta_3 \text{mar}_{it} + \beta_4 \text{yngkid}_{it} + \beta_5 \text{gdhlth}_{it} + a_i + u_{it},$$

$$t = 1, 2.$$

• ماذا يمكن أن يمثل  $a_i$  من عوامل (ثابتة)؟

الأثر الثابت أو الأثر غير المشاهد  $a_{it}$  يمثل الفروق الفردية (بعض الأفراد أكثر حيوية أو أكثر طاقة،...).

• هل من المهم احتساب العوامل المتضمنة في  $a_i$ ؟ لم؟

من المهم احتساب هذه العوامل لأنها من المرجح مرتبطة بالمفسرة  $\text{totwrk}$ ، فالأفراد ذوي طاقة عالية قد يكونون أكثر نشاطا إضافة إلى كونهم أقل حاجة للنوم من غيرهم ...

• لماذا لم ندرج في النموذج متغيرة الجنس أو العرق؟

على عكس البيانات المقطعية المجمعة، فإن مثل هذه المتغيرات لا تدرج لأنها لا تتغير مع الزمن (لو أدرجت يأتي الفارق معدوم لجميع المفردات)، فهي مدمجة في  $a_i$ .

لتقدير الفرق بين السنتين، نقدر الدالة التالية:

$$\Delta \text{snap}_i = \delta_0 + \beta_1 \Delta \text{totwrk}_i + \beta_2 \Delta \text{educ}_i + \beta_3 \Delta \text{mar}_i + \beta_4 \Delta \text{yngkid}_i + \beta_5 \Delta \text{gdhlth}_i + \Delta u_i,$$

ما هو الافتراض الأساسي للحصول على مقدرات متقاربة بطريقة المربعات الصغرى؟

بافتراض أن  $\Delta u_{it}$  ليست مرتبطة بتغير أي من المفسرات، يمكن الحصول على مقدرات متقاربة بطريقة المربعات الصغرى.

أدخلنا البيانات فكانت النتيجة كالتالي:

$$\Delta \ln \text{ap}^{\wedge} = -92.63 - 0.227 \Delta \text{totwrk} - 0.024 \Delta \text{educ} + 104.21 \Delta \text{mar} + 94.67 \Delta \text{yngkid} + 87.58 \Delta \text{gdhlth}$$

(45.87) (0.036) (48.759) (92.86) (87.65) (76.60)

N= 239, R<sup>2</sup> = 0.15.

• ما هو مدلول معامل **totwork**؟

ساعة عمل إضافية تقابلها خسارة بـ 0.227 ساعة أي 13 دقيقة نوم: (0.227 x 60 = 13.62 min)

• هل التأثير مؤكد لمدة العمل؟

إحصائية الاختبار t دالة جدا فالتأثير مؤكد: -0.227/0.036 = -6.31 .

• هل الثابت والمتغيرات الأخرى دالة؟

كل المتغيرات الأخرى غير دالة ما عدا الثابت.

• اختبار F لمجمل المتغيرات الأخرى ما عدا **totwrk** أعطى مستوى معنوية **0,49**. ماذا يعني ذلك؟

يعني أن المتغيرات الأخرى غير ماثرة معا ويمكن أن تسحب من الدالة.

• الانحراف المعياري ل **educ** هل يظهر لك كبيرا؟ ماذا يعني ذلك؟ وما السبب المحتمل؟

نعم الخطأ المعياري كبير لهذه المتغيرة مقارنة بقيمة المعامل، والسبب هو قلة تباين **educ**.

في الواقع فإن نسبة كبيرة من العمال المستجوبين (76.6%) لم يتغير عدد سنوات دراستهم خلال الفترة المدروسة، و90% زاد عدد سنوات دراستهم بـ 1 سنة على الأكثر. هذا يعني أساسا أن مع هذا الضعف في تباين المتغيرة يصعب تقدير معاملها بدقة. في جميع الأحوال، فإن قيمة المعامل نفسها صغيرة.

## 2-5. النماذج ذات التأخرات

بيانات البانل يمكن أن تستخدم لتقدير نماذج بتأخرات سُلمية منتهية *modèles à retards échelonnés finis* أي نماذج يمتغيرات مفسرة متأخرة (أي أن المتغيرة التابعة سجلت في تاريخ لاحق وليس في نفس تاريخ تسجيل مشاهدات المتغيرات المستقلة). المثال التالي يوضح ذلك.

**مثال الجريمة وحل القضايا.** إستخدم (Eid 1994) بيانات الشرطة في النرويج، لتقدير نموذج بتأخرات سُلمية لمعدل الجريمة. المتغيرة المفسرة هي (**clrprc**) نسبة قضايا الجرائم والجنح المحلولة وتم تسجيلها للسنة السابقة والتي قبلها. لغرتم معدل الجريمة (**crime**) حسب في 1972 و1978. حسب الباحث فإن الاجتهاد في حل القضايا لسنة ما له أثر على معدل الجريمة في السنوات المقبلة.

• أكتب نموذج الأثر الفردي:

$$\text{Log}(\text{crime})_{it} = \beta_0 + \delta_0 d78_t + \beta_1 \text{clrprc}_{it-1} + \beta_2 \text{clrprc}_{it-2} + a_i + u_{it}$$

• بفرض أننا استخدمنا دالة الفروق الأولى، فجاءت الدالة كما يلي:

$$\Delta \text{Log}(\text{crime})^{\wedge} = 0.86 - 0.004 \Delta \text{clrprc}_{-1} - 0.0132 \Delta \text{clrprc}_{-2}$$

(0.064) (0.0047) (0.0052)

N= 53, R<sup>2</sup> = 0.193, R<sup>2</sup>aj = 0.161

- هل المتغيرات دالة؟ فسر الأثر.
- التأخر الثاني دال، وإشارته سالبة، مما يعني أن نسبة الجريمة حالياً تتخفف إذا كانت هناك زيادة في السنة ما قبل الماضية في نسبة القضايا المحلولة.
- بالتحديد، زيادة في clrpc ب 10 نقاط مئوية، تؤدي إلى تقليل نسبة الجريمة بما يقارب 13.2% سنتين بعد ذلك.

## 2-6. تطبيق دالة الفروق الأولى على تقييم السياسات

بيانات البنابل هي أداة قوية لتحليل وقياس أثر سياسة أو برنامج أو حدث اقتصادي ما. نأخذ عينة من المفردات قبل المعالجة، جزء منها يخضع للمعالجة فيسمى عينة معالجة والباقي يسمى عينة مراقبة، لكن الفرق عما رأيناه في البيانات المقطعية المجمعة هو أن مفردات العينة قبل المعالجة هي نفسها مفردات العينة بعد المعالجة.

### مثال: برنامج التكوين

بفرض أننا نريد تقييم مردودية برنامج لتكوين مهني للعاملين لفائدة مؤسسات من أجل تقليل التالف scrap. المفسرة هي grant وهي متغيرة ثنائية تأخذ 1 للمؤسسات المستفيدة من منحة التكوين، و 0 لغيرها. البيانات مأخوذة عن عينة من المؤسسات في 1987 و 1988.

- أكتب نموذج الأثر الثابت.

$$\text{Scrap}_{it} = \beta_0 + \delta_0 y88 + \beta_1 \text{grant}_{it} + a_i + u_{it}, t = 1, 2.$$

- ماذا يمكن أن يتضمن  $a_i$  من خصائص؟ وهل يمكن اعتبارها ثابتة في الزمن؟
- يتضمن  $a_i$  الفروقات بين المؤسسات، مثل كفاءة العامل المتوسط، رأس المال، وكفاءة التسيير، وهي متغيرات يمكن اعتبارها ثابتة لأن الفترة قصيرة سنتين (لكن ليس لو كانت الفترة طويلة).
- هل يجب أن نخشى من ارتباط  $a_i$  بالمتغيرة المستقلة؟
- يمكن أن تكون  $a_i$  مرتبطة بالمتغيرة المفسرة grant، بمعنى البرنامج الحكومي يعطي الأولوية للمؤسسات الأعلى كفاءة مثلاً في المنحة من أجل إنجاز المشروع مثلاً. يمكن أن يحدث العكس أيضاً، بحيث تدعم المؤسسات الأدنى مستوى. في الحالتين فإن استخدام بيانات مقطعية أو مجمعة يعطي مقدرات متحيزة وغير متقاربة.
- كيف يمكن أن نحذف  $a_i$  من النموذج؟ أكتب الدالة وفسر المتغيرة المفسرة.
- بتحويل الدالة إلى دالة فروق أولية نحذف  $a_i$  ونحصل على:

$$\Delta \text{Scrap}_i = \delta_0 + \beta_1 \Delta \text{grant}_i + \Delta u_i$$

grant هي متغيرة ثنائية تأخذ 1 للمؤسسات المستفيدة من منحة، وبما أنه في 87 كل المؤسسات لم تكن قد حصلت على المنحة، فإن :

$$\Delta \text{grant}_i = \text{grant}_{i2} - \text{grant}_{i1} = \text{grant}_{i2}$$

أي أن  $\Delta \text{grant}_i$  هي متغيرة ثنائية تعين المؤسسات المستفيدة من المنحة من غيرها.

- إذا أعطى تقدير الدالة النتيجة التالية، فسر معامل الانحدار في الدالة.

$$\Delta \text{Scrap}^{\wedge} = -0.564 - 0.739 \Delta \text{grant}$$

$$(0.405) \quad (0.683)$$

$$N = 54, R^2 = 0.022.$$

المعامل  $\beta_1$  يعني أن المؤسسات المستفيدة من المنحة يقل فيها معدل التالف في المتوسط ب 0.739. غير أن هذا المعامل غير دال، فالعلاقة غير مثبتة. الثابت هو الآخر لا يختلف بدلالة عن الصفر.

• نحصل على نتائج أفضل بتحويل التابع:

$$\Delta \log(\text{Scrap}^i) = -0.057 - 0.317 \Delta \text{grant}^i$$

$$(0.097) \quad (0.164)$$

$$N = 54, R^2 = 0.067.$$

فسر المعامل -0.317

حسب معامل الانحدار الجزئي، الاستفادة من المنحة تقلص نسبة التالف بحوالي 31 بالمائة، بالتحديد 27.2% (  $\exp(-0.317) - 1 = -0.272$  )  
لكن المعامل غير دال، لذلك لا يمكن الاستدلال على وجود علاقة.

بصفة عامة، إذا رمزنا للمتغيرة المفسرة، المشاركة في برنامج ما، بمتغيرة ثنائية prog، فإن النموذج الأبسط للأثر الثابت يكتب كما يلي:

$$Y_{it} = \beta_0 + \delta_0 d_{2i} + \beta_1 \text{prog}_{it} + a_i + u_{it}, t = 1, 2.$$

في حالة كون المساهمة في البرنامج تتم في الفترة الثانية فقط، فإن المقدر بالمربعات الصغرى للمعامل  $\beta_1$  في دالة الفروق يمكن حسابه كما يلي:

$$\hat{\beta}_1 = M \Delta y_{\text{Traitement}} - M \Delta y_{\text{contrôle}}$$

أي متوسط فروق (بين الفترتين) عينة المعالجة مطروحا منه متوسط فروق عينة المراقبة.

ميزة هذه الطريقة في بيانات البانل عن مثيلتها في البيانات المجمعة، أننا هنا نحسب الفروق لنفس المفردات، وبالتالي نأخذ بعين الاعتبار الفروق الفردية.

في حالة كون المساهمة في البرنامج ليست مقصورة على الفترة الثانية، عندها لا يمكن حساب المقدر أعلاه بنفس الطريقة، لكن التفسير يبقى هو التغير في متوسط y العائد للمساهمة في البرنامج.

**الخلاصة** هي أن استخدام بيانات البانل يسمح بأخذ في الحسبان العوامل التي تتغير في الزمن والتي يمكن أن تكون مرتبطة بخصائص البرنامج.

هذه الطريقة للفروق الأولى تصلح أيضا لتحليل آثار سياسة تتغير بحسب البلد أو المدينة، وهو ما يشرحه المثال التالي.

### مثال 2 قوانين مكافحة القيادة في حالة سكر

لتقليل حوادث السير تتبع الولايات المتحدة سياسات مختلفة لمكافحة القيادة في حالة سكر، ندرس هنا سياستين: منع الراكبين من حيازة عبوات مفتوحة open، وسياسة السحب الإداري الفوري لرخصة السياقة admn. لتحليل أثر هاتين السياستين على عدد الحوادث يمكن استخدام بيانات مقطعية لمجموعة من الولايات، التابع هو عدد الحوادث والمفسرات هي متغيرات ثنائية تعبر عن السياسة المتبعة من قبل الولاية. العائق في هذه الطريقة هو أن وجود قانون ما قد يكون مرتبطا بعدد الحوادث في الولاية في السنوات الأخيرة، خاصة المرتبطة بالكحول.

الطريقة الأفضل هي استخدام بيانات البانل لفترة تبنت فيها بعض الولايات سياسة مختلفة. الدراسة التي نحن بصددنا تتضمن بيانات ل 50 ولاية زائد مقاطعة كولومبيا، بين 1985 و 1990. المتغيرة التابعة هي عدد الوفيات بسبب حوادث السير في 100 مليون ميل بالسيارة  $d_{thrt}$ . في 1985 كانت 19 ولاية تعتمد سياسة منع العبوة المفتوحة مقابل 22 في 1990، وفي 1985، 21 ولاية اعتمدت سياسة سحب رخصة السياقة مقابل 29 في 1990.

تقدير النموذج بطريقة الفروق الأولى أعطى النتيجة التالية:

$$\Delta d_{thrt} = -0.497 - 0.42 \Delta open - 0.151 \Delta admn$$

(0.052) (0.206) (0.117)

$$N = 51, R^2 = 0.119$$

- فسر إشارة وقيم معاملات الانحدار الجزئية.
- استخدام سياسة العبوة قلص معدل الحوادث ب 0.42، والعلاقة دالة عند مستوى معنوية 5 بالمائة. سياسة سحب الرخصة كان لها أثر أقل، وهو أثر غير دال ( $t = -1.29$ ).
- فسر إشارة وقيمة الثابت.
- الثابت يدل أن معدل الحوادث هو في تناقص بين السنتين 85 و 89 في مجمل الولايات، سواء حدث تغيير في السياسة أم لا. الولايات التي تبنت سياسة العبوة المفتوحة عرفت انخفاضا أكبر في عدد الحوادث.
- بالنسبة لولاية واشنطن  $\Delta admn = -1$  فسر ما معنى ذلك.
- بما أن المتغيرة هي حاصل الفرق بين قيمة 1990 و 1985، فهي تساوي -1 عندما تكون القيمة في 1990 هي 0 والقيمة في 1985 هي 1، أي أن الولاية كانت تستخدم سياسة سحب رخصة السياقة في 1985 ثم تراجع عنها في 1990.

## 7-2. دالة الفروق الأولى لأكثر من فترتين

رأينا أنه في حالة وجود فترتين فقط يكتب نموذج الاثار الفردية كما يلي:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + \alpha_i + u_{it}, t = 1, 2.$$

وأن دالة الفروق الأولى تأتي (من خلال طرح دالة الفترة الأولى من الثانية) كما يلي:

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_{i1} + \beta_2 \Delta x_{i2} + \dots + \Delta u_i$$

(يزول  $\alpha_i$ ، والثابت  $\beta_0$  بالطرح، ويزول  $d_{2t}$  لأنه في الفترة الأولى يساوي الصفر وفي الثانية يساوي 1). حجم العينة هنا هو N وليس NT.

مما سبق، في حالة تعدد الفترات نضيف متغيرات زمنية، مثلا في حالة  $T = 3$ ، نضيف متغيرة تعبر عن السنة الثالثة. من المناسب أيضا أن نكيف الترميز لتحسين مقروئية المعاملات، ونكتب نموذج الأثر الثابت كما يلي:

$$y_{it} = \delta_1 + \delta_2 d_{2t} + \delta_3 d_{3t} + \beta_1 x_{it1} + \dots + \beta_j x_{itj} + \dots + \beta_k x_{itk} + a_i + u_{it}, t = 1, 2, 3.$$

عدد المشاهدات الإجمالي هو 3N.

المتغيرتين  $d_2$  و  $d_3$  ثنائيتين. السنة الأولى هي سنة مرجعية فلا تظهر لها متغيرة تمثلها. بهذا الشكل يصبح الثابت للسنة<sup>1</sup> الأولى:  $\delta_1$ ، للسنة الثانية هو:  $\delta_2 + \delta_1$ ، وللسنة الثالثة هو:  $\delta_3 + \delta_1$ .

**لاحظ.** ما يعنينا بالدرجة الأولى هو معاملات الانحدار الجزئية. مرة أخرى إذا كانت الفروقات الفردية  $a_i$  مرتبطة بالمتغيرات المفسرة، فإن التقدير باستخدام المربعات الصغرى ببيانات مجمعة يكون غير متقارب ومنحاز. الفرضية الأساسية هي أن الخطأ البسيط أي  $u_{it}$  غير مرتبط بالمتغيرات المفسرة في مختلف التواريخ، أي:

$$\text{Cov}(x_{itj}; u_{is}) = 0, \text{ for all } t, s, \text{ and } j.$$

هذا يعني أن المتغيرات المفسرة هي خارجية تماماً<sup>2</sup> عندما نحذف الأثر الثابت  $a_i$ . عادة تكون الفرضية غير محققة في حالة إهمال متغيرة مهمة أو متغيرات في الزمن، أو إذا كانت المتغيرات لم يتم قياسها بطريقة صحيحة. إذا كانت  $a_i$  مرتبطة ب  $x_{itj}$  عندها تكون  $x_{itj}$  مرتبطة بالخطأ المركب  $v_{it} = a_i + u_{it}$ .

للتخلص من  $a_i$  في حالة  $T = 3$ ، نطرح (في المعادلة أعلاه) الفترة الأولى من الثانية، ونطرح الثانية من الثالثة، فتأتي الدالة كما يلي:

$$\Delta y_{it} = \delta_2 d_{2t} + \delta_3 \Delta d_{3t} + \beta_1 \Delta x_{it1} + \dots + \beta_k \Delta x_{itk} + \Delta u_{it}, t = 2, 3.$$

تعطي المعادلة هنا فترتين لكل مفردة:  $t = 2, 3$ ، لا توجد دالة فروق ل  $t=1$  لأنه ليس هناك ما يطرح منها<sup>3</sup>.

إذا كانت شروط النموذج الخطي التقليدي محققة فإن المربعات الصغرى على بيانات مجمعة تعطي مقدرات غير منحازة، والإحصائيات  $t$  و  $F$  صحيحة. الشرط الأساسي هو عدم الارتباط بين  $\Delta u_{it}$  و  $x_{itj}$  من أجل أي قيمة ل  $j$  من أجل  $t = 2, 3$ .

كون الدالة أعلاه لا تتضمن ثابتاً يسبب مشكلة في حساب  $R^2$ .

$$\text{For } t = 2 : \Delta d_{2t} = 1; \Delta d_{3t} = 0;$$

$$\text{For } t = 3 : \Delta d_{2t} = -1; \Delta d_{3t} = 1$$

لذلك يستحسن تقدير دالة الفروق الأولى بثابت ومتغيرة واحدة فقط للزمن، عادة الثالثة. في هذه الحالة تصبح الدالة كما يلي:

$$\Delta y_{it} = \alpha_0 + \alpha_3 d_{3t} + \beta_1 \Delta x_{it1} + \dots + \beta_k \Delta x_{itk} + \Delta u_{it}$$

الطريقة هي نفسها من أجل أكثر من ثلاث فترات، حيث نضيف ثنائية متغيرات تعبر عن الفترات الموالية. في حالة توفر نفس العدد من الفترات  $T$  لكل مفردة، عندها نقول أن بيانات البائل أسطوانية **panel cylindré**. في حالة  $T$  صغيرة مقارنة مع  $N$  يتعين أن ندرج متغيرة لكل فترة زمنية من أجل احتساب التغيرات التي لم يتم نمذجتها. بهذه الطريقة يكون لدينا دالة تظهر كما يلي:

$$\Delta y_{it} = \alpha_0 + \alpha_3 d_{3t} + \alpha_4 d_{4t} + \dots + \alpha_T d_{Tt} + \beta_1 \Delta x_{it1} + \dots + \beta_k \Delta x_{itk} + \Delta u_{it}, t = 2, 3, \dots, T.$$

<sup>1</sup> كان يمكن استخدام  $\beta_0$ ، كما في ترميز حالة الفترتين، تغيير الترميز هدفه الانسجام مع معاملات الزمن للسنوات الأخرى لا غير.

<sup>2</sup> تلغي هذه الفرضية الحالات التي تكون فيها المفسرة تتأثر بالتغير في  $u_{it}$  مثلما يحدث عندما تكون المتغيرة المفسرة هي متغيرة تأخر.

<sup>3</sup> يحذف الثابت  $\delta_1$  بفعل الطرح. ويستبدل  $d_{3t}$  ب  $\Delta d_{3t}$  لأننا نطرح السنة السابقة، وهي هنا السنة الثانية، بينما يبقى  $d_{2t}$  على حاله لأنه يمثل الفرق عن السنة السابقة له مباشرة.

عدد المشاهدات في هذه الحالة هو  $N(T-1)$ ، لأن السنة الأولى ليس عنك ما يطرح منها. من السهل تقدير الدالة أعلاه بطريقة MCO على بيانات مجمعة، بشرط أن تكون البيانات تم ترتيبها بعناية والفروق تم حسابها بعناية<sup>1</sup>. في المثال التالي نرى أهمية الفروق الأولى في حالة بيانات البائل لأكثر من فترتين في تقييم السياسات العمومية.

**مثال 1: أثر سياسة المناطق الحرة على البطالة** درس بابك (1994) Papke أثر برنامج لإنشاء "مناطق مؤسسات" (ez) في مدن ولاية إنديانا في الولايات المتحدة على البطالة، هذه الأخيرة تم قياسها بعدد طلبات منحة البطالة (unemployment claims: uclams). تم تحليل 22 مدينة في الولاية في كل سنة خلال ثمن سنوات من 80 إلى 88. في هذه الفترة تم تعيين 6 مناطق مؤسسات في سنة 84، و4 مناطق أخرى في 85. بينمت 12 مدينة لم تعين كمناطق مؤسسات خلال الفترة وهي تشكل عينة المراقبة.

• **كم هو إذن الحجم الإجمالي للعينة؟**

عدد السنوات هو  $8-1=9$ ، وعدد المدن 22 إذن الحجم الإجمالي للعينة هو  $N(T-1) = 22(8) = 176$ . لتقييم السياسة يمكن وضع النموذج البسيط التالي:

$$\text{Log}(uclams_{it}) = \theta_t + \beta_1 ez_{it} + a_i + u_{it}$$

المتغيرة  $ez_{it}$  هي متغيرة ثنائية تأخذ 1 للمدن  $i$  التي لها منطقة مؤسسات في اللحظة  $t$ .

• **حجم البطالة تقلص خلال الفترة في عموم مدن الولاية، كيف يظهر ذلك في الدالة.**

في قيم سالبة ل الثابت في كل فترة  $\theta_t$

• **هل هناك ما يدعو للاعتقاد بوجود ارتباط بين  $ez_{it}$  و  $a_i$ ؟**

الأثر الثابت  $a_i$  يمثل الفروقات الفردية بين المدن (مناخ الأعمال...)، لكي يكون عدم الارتباط بين  $ez$  و  $a_i$  يجب أن تعين المدن للبرنامج عشوائيا، وبما أن هذا لم يتم عشوائيا فمن المحتمل أن المدن الأضعف في مناخ الأعمال (بطالة أعلى) هي التي منحت مناطق مؤسسات، وبالتالي وجود ارتباط موجب بين  $ez$  و  $a_i$ .

• **كيف يمكن معالجة الارتباط المحتمل بين  $ez_{it}$  و  $a_i$  وكيف تصبح الدالة (مع الاحتفاظ بثابت)؟**

للتخلص من  $a_i$  نستخدم دالة الفروق الأولى:

$$\Delta \text{Log}(uclams_{it}) = \alpha_0 + \alpha_1 d82_t + \dots + \alpha_7 d88_t + \beta_1 ez_{it} + \Delta u_{it}$$

• **ماذا تمثل المتغيرة التابعة؟**

المتغيرة  $\Delta \text{Log}(uclams_{it})$  تمثل تقريبا المعدل السنوي للنمو في الطلب على منح البطالة خلال الفترة من  $t-1$  إلى  $t$ .

• **التقدير ل  $\beta_1$  أعطى: -0.182، بانحراف معياري 0.078. ماذا يعني ذلك بالتحديد؟**

وجود منطقة مؤسسات في المدينة في الفترة  $t$  يقلل الطلب على منح البطالة. العلاقة دالة لأن  $t = -2.33$ . بالتحديد، وجود منطقة مؤسسات في مدينة ما يقلل الطلبة على منح البطالة بنسبة 18.2% تقريبا. القيمة الحقيقية للتخفيض هي:  $\exp(-0.182) - 1 = -16.64$

• **إحصائية F لاختبار عدم ثبات تباين الخطأ من أجل مختلف المفردات  $\text{hétéroscédasticité}$**

ليراوش-باقان Breuch-Pagan جاءت:  $\text{sig.} = 0.557$  ;  $F = 0.85$  ماذا يعني ذلك؟

أي ليس هناك دليل على عدم اختلاف التباين.

<sup>1</sup> عدد الأسطر هو NT، حيث الأسطر الأولى (T سطر) تمثل مشاهدات المفردة الأولى، مرتبة زمنيا، وهكذا، ثم تحسب الفروق ويرمز الانتقال من  $(t-1)$  إلى  $t$  في السطر الموافق للفترة رقم  $t$ . بهذا، من أجل  $t=1$ ، يتعين أن تعطي قيما مفقودة للمشاهدات  $N$ .



في هذا المثال كانت لدينا مفسرة واحدة؛ يمكن بسهولة إدراج عدة متغيرات مفسرة في النموذج. هذا التعدد لا يغير في شيء تفسير المعاملات. يمكن أيضا تحويل المتغيرات إلى اللغز، ويفسر بنفس الطريقة كما في البيانات المقطعية. في حالة وجود اختلاف للتباين هناك طرق لتصحيح الخطأ المعياري (الحصول على خطأ معياري مقاوم) والارتباط الزمني (hétéroscédasticité et autocorréation sérielle). في المثال التالي وجد أن اختبار اختلاف التباين دال جدا (اختبار وايت White test of heteroscedasticity).

### مثال 2: معدل الجريمة

استخدم كورنول وترومبل (Kornwell and Trumbull (1994) بيانات 90 مقاطعة في ولاية كارولينا الشمالية، من سنة 1981 إلى 1987، لتقدير نموذج أثر ثابت لمعدل الجريمة. يمكن إدراج عدة متغيرات ضمن  $a_i$  مثل الموقع الجغرافي، الموقف من الجريمة، التوجه التاريخي، الاتفاقيات لتحويل المجرمين ... معدل الجريمة يقصد به عدد الجرائم للفرد الواحد. المتغيرات هي:

$prbarr$  يمثل احتمال الاعتقال،  $prbconv$  احتمال الإدانة (في اعتقال ما)، و  $prbpris$  هي احتمال أن تشمل العقوبة فترة من السجن،  $avgsnt$  هي متوسط مدة العقوبات المنصوصة،  $polpc$  عدد ضباط الشرطة للسكان. مثل دراسات الجريمة في الاقتصاد القياسي نستخدم اللغز لكل المتغيرات لقياس المرونة. ندرج أيضا عدد من متغيرات مؤشرة على الزمن لاحتمال التوجه العام للظاهرة في الولاية. يمكننا استخدام البيانات من 1982 إلى 1987 لتقدير دالة الفروق.

$$\Delta \log(\text{cmrte}^a) = 0.008 - 0.1d83 - 0.048d84 - 0.005d85 + 0.028d86 + 0.041d87 \\ - 0.327\Delta \log(\text{prbarr}) - 0.238\Delta \log(\text{prbconv}) - 0.165\Delta \log(\text{prbpris}) - 0.022\Delta \log(\text{avgsen}) + 0.398 \Delta \log(\text{polpc}) \\ N = 540, R^2 = 0.433, R^2_{aj} = 0.422$$

#### • فسر معامل $prbarr$

زيادة في احتمال التوقيف ب 10% تؤدي لتخفيض نسبة الجريمة ب 3.27 بالمائة تقريبا.

#### • فسر معامل $polpc$

إشارة معامل عدد رجال الشرطة للسكان موجبة، وهذا لو اعتمدنا علاقة سببية يعني أن زيادة نسبة عدد رجال الشرطة للسكان ب 1% يؤدي لزيادة معدل الجريمة بحوالي 0.4%، لكن الواقع أن العلاقة هي في الاتجاه المعاكس، المدن التي تزيد فيها الجريمة توظف أعداد إضافية من رجال الأمن، أي أن المتغيرة  $polpc$  هي متغيرة داخلية للنموذج  $endogène$ . المعامل إذن لا يمكن أن يقرأ بالطريقة السببية التقليدية المذكورة. هناك طرق لمعالجة مثل هذه المتغيرات. سبب آخر يمكن أن عدد القضايا ( $y$ ) مرتبط بالتصريح، ولذلك كل ما كان عدد رجال الشرطة أكبر زاد عدد القضايا المصرح بها.

## فصل 3. طريقتي الأثر الثابت والأثر العشوائي

### طريقة الأثر الثابت بالتحويل الداخلي - طريقة الأثر العشوائي

بعد أن تطرقنا لنموذج الأثر الفردي وتقديره بطريقة دالة الفروق الأولى، نتطرق لطريقة الأثر الثابت من خلال التحويل الداخلي وطريقة الأثر العشوائي.

### 1. طريقة الأثر الثابت (Fixed Effects) بالتحويل الداخلي *within*

الدالة

المقارنة بين FE و FD

#### 1-1. الدالة

دالة الفروق الأولى هي طريقة من بين طرق تقدير نموذج الأثر الفردي غير المشاهد  $a_i$ . هناك طريقة أخرى في التقدير تعطي نتائج أفضل في ظل فرضيات معينة تدعى طريقة التحويل الداخلي. لنأخذ نموذجا بمفسرة واحدة. من أجل كل مفردة  $i$  لدينا:

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \dots, T$$

لنأخذ لكل مفردة متوسط قيم الدالة في الزمن:

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i$$

With:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it};$$

$$\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it};$$

$$\bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it};$$

للتخلص من  $a_i$  نطرح المتوسط من الدالة، أي نحول إلى المتغيرة الممركزة، فنجد:

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + (a_i - a_i) + (u_{it} - \bar{u}_i)$$

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i); \quad t = 1, 2, \dots, T$$

أحيانا نكتب أيضا (اختصارا):

$$y_{it}^{\cdot} = \beta_1 x_{it}^{\cdot} + u_{it}^{\cdot}; \quad t = 1, 2, \dots, T$$

الآن وقد تخلصنا من الأثر الثابت  $a_i$  يمكن التقدير بطريقة المربعات الصغرى على البيانات الممركزة (time demeaned data). حجم العينة هو  $NT$ .

**لاحظ:**

بعض البرمجيات تضيف المتوسط العام إلى طرفي المعادلة، فتضيف المتوسط العام ل  $y$  في الطرف الأيسر، وتضيف متوسط المفسرة في الطرف الأيمن.

$$y_{it} - \bar{y}_i + \bar{y} = \beta_1(x_{it} - \bar{x}_i + \bar{x}) + (u_{it} - \bar{u}_i + \bar{u}); t = 1, 2, \dots, T$$

إذا تم إدراج ثابت في النموذج  $\beta_0$  فإن قيمته تمثل متوسط الأثر الثابت.

$$y_{it} - \bar{y}_i = \beta_0 + \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i); t = 1, 2, \dots, T$$

المقدر بهذه الطريقة يدعى مقدر الأثر الثابت أو المقدر الداخلي within estimator. سبب التسمية هو أن طريقة المربعات الصغرى تعتمد في هذه الحالة على التباين الداخلي للمفردة في الزمن. هناك أيضا طريقة التقدير البيئي between estimator لكنها منحازة في حالة وجود ارتباط بين  $a_i$  و  $x_i$  من السهل إدراج متغيرات مفسرة أخرى، ويكون ذلك بنفس الطريقة، أي التحويل إلى المركزة. نموذج الأثر الثابت في صيغته الأصلية (بدون ثابت) هو:

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

من خلال طرح المتوسط يصبح النموذج كما يلي:

$$y_{it} - \bar{y}_i = \beta_1(x_{it1} - \bar{x}_{i1}) + \beta_2(x_{it2} - \bar{x}_{i2}) + \dots + \beta_k(x_{itk} - \bar{x}_{ik}) + u_{it} - \bar{u}_i, t = 1, 2, \dots, T$$

ويتم تقديره بطريقة المربعات الصغرى على البيانات المجمع.

في ظل فرضية الخارجية التامة **stricte exogeneity** للمتغيرات المفسرة، يكون مقدر الأثر الثابت غير متحيز: بصيغة تقريبية يعني ذلك أن الخطأ الذاتي  $u_{it}$  غير مرتبط بالمتغيرات المفسرة في أي فترة.

كما في طريقة الفروق الأولى، يسمح مقدر الأثر الثابت بارتباط  $a_i$  بالمتغيرات المفسرة في أي فترة.

من أجل تطبيق جيد لطريقة المربعات الصغرى يفترض التجانس homoscedasticity وعدم الارتباط الزمني للخطأ  $u_{it}$ .

عيب أساسي في هذه الطريقة أنها تلغي كل متغيرة غير متغيرة في الزمن فلا يعرف معاملها، لأن  $x_{it} = 0$  من أجل أي قيمة ل  $it$  إذا كانت المتغيرة ثابتة لا تتغير، لذلك لا يمكن إدراج متغيرات ثابتة في الزمن مثل الجنس، مساحة المدينة ...

إذا كان الغرض هو تقدير تأثير المتغيرات غير المتغيرة في الزمن فيتعين اختيار طريقة أخرى (POLS أو Between estimators).

(نستخدم طريقة الأثر الثابت عندما نرجح وجود فروقات فردية (تأثر على التابع غير مدرجة كمفسرات) وهذه الفروقات الفردية مرتبطة بالمفسرات المدرجة.)

**2-1. المقارنة بين طريقة FE و FD**

- الطريقتان تعطيان نفس التقدير عندما  $T = 2$ ، لكن عندما  $T > 2$  تعطي الطريقتان نتائج مختلفة
- في حالة  $u_{it}$  غير مرتبطة، FE أكثر فعالية من FD
- في حالة  $\Delta u_{it}$  غير مرتبطة ذاتيا، FD أفضل، لذلك يجب دائما اختبار وجود ارتباط ذاتي في  $\Delta u_{it}$
- يستحسن تجريب الطريقتين، إذا كانت النتائج لا تختلف كثيرا فذلك دليل جيد.

## 2. طريقة الأثر العشوائي

الدالة

اختبار هاوسمن

### 1-2. دالة الأثر العشوائي

كما في طريقة POLS فإن طريقة الأثر العشوائي (RE) تضع  $\alpha_i$  ضمن حد الخطأ، أن لكل مفردة حد خطأ مركب

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

طريقة الأثر العشوائي تضع افتراضات أكثر من POLS. نفترض أن المعامل  $\alpha_i$  غير مرتبط بالمفسرات، فلا نحتاج للتخلص منه، ونفترض أنه عشوائي، وبالتالي ينظم إلى الخطأ ولا نحتاج لتقديره.

النموذج:

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 x_{it} + (\alpha_i + u_{it}), \quad t = 1, 2.$$

لاحظ إمكانية إدراج في النموذج متغيرة ثنائية للزمن.

نعبر عن افتراض الاستقلال كما يلي<sup>1</sup>:

$$\text{Cov}(\alpha_i ; x_{it}) = 0$$

في هذه الحالة كلا النموذجين RE و FE يعطيان مقدرات متقاربة consistent :

$$B^{\wedge}_{RE} \text{ consistent}, B^{\wedge}_{FE} \text{ consistent}$$

لكن مقدرات RE أكثر دقة: أي أن خطأها المعياري أقل:

$$SE(B^{\wedge}_{RE}) < SE(B^{\wedge}_{FE})$$

• في حالة عدم تحقق هذه الفرضية، طريقة RE تكون غير متقاربة بينما تبقى طريقة FE متقاربة.

لتحقيق هذه الفرضية نحتاج عادة لإدراج عدة متغيرات مفسرة في النموذج بحيث تستوعب مجمل التباين في التابع بما لا يدع مجالاً تقريباً للفروقات الفردية.

لاختبار فرضية عدم الارتباط هذه (للمفاضلة بين النموذجين: RE و FE) يستخدم اختبار هاوسمن Hausman.

$$\text{var}(v_{it}) = \text{var}(\alpha_i) + \text{var}(u_{it}) = 0$$

في الواقع الافتراض المطلوب هو الخارجية التامة والاستقلالية strict exogeneity and orthogonality between  $x_{it}$  and  $\alpha_i$

التعبير عن الخارجية التامة يمكن أن يكون كما يلي:

<sup>1</sup> يمكن أيضاً كتابة افتراض الخارجية التامة بدلالة الخطأ البسيط (رغم أن الكتابة أعلاه أوضح) وكتابة افتراض العمودية كما يلي:

assumption RE.1:

(a)  $E(u_{it} | x_i; \alpha_i) = 0, t : 1; \dots ; T.$

(b)  $E(\alpha_i | x_i) = E(\alpha_i) = 0$

$$E(y_{it} | x_{i1}; x_{i2}; \dots ; x_{iT}; \alpha_i) = E(y_{it} | x_{it}; \alpha_i) = \beta x_{it} + \alpha_i$$

ويعني (المساواة الأولى هي المهمة) عدم ارتباط  $y_{it}$  في فترة  $t$  بالمدخلات  $x_i$  في فترات سابقة. مثلاً: إذا كانت  $y_{it}$  هي محصول الصويا لمزرعة ما  $i$ ، والمدخلات هي راسالمال، كمية السماد، كمية العمل، كمية الأمطار، ومتغيرات أخرى مشاهدة، فإن  $\alpha_i$  يمكن أن يلتقط فروقات فردية غير مشاهدة وغير متغيرة مع الزمن مثل خصوبة الأرض، مهارة التسيير، وغيرها. الافتراض الطبيعي هو أنه طالما تم احتساب المتغيرات المشاهدة مع الأثر غير المشاهد فإن مدخلات سنوات أخرى ليس له تأثير على المحصول لهذه السنة، وهذا لأننا احتسبنا الفروقات الفردية. أنظر Woldridge, p253

## 2-2. اختبار هاوسمن

اختبار هاوسمن للاستقلال بين الفروقات الفردية والمفسرات، إذا جاء الاختبار دالاً فهذا يعني الذهاب إلى طريقة الأثر الثابت وليس الأثر العشوائي لأن هناك ارتباط ما بين الفروقات الفردية والمفسرات. كيف ذلك؟  
الفرضية الصفرية لهذا الاختبار هي أن نموذج الأثر العشوائي ملائم (consistent)، والفرضية البديلة هي أن نموذج الأثر الثابت هو الملائم، أو بعبارة أخرى:

$$H_0 : \text{Cov}(\alpha_i ; x_{it}) = 0$$

في هذه الحالة كلا النموذجين RE و FE يعطيان مقدرات متقاربة consistent.

في ظل الفرضية الصفرية، تتبع إحصائية اختبار هاوسمن توزيع ك2 بدرجة حرية واحدة (يشبه شكل التوزيع الأسّي)، وتحسب الإحصائية كما يلي:

$$W = (B^{FE} - B^{RE})^2 / (\text{var}(B^{FE}) - \text{var}(B^{RE}))^2 \sim X^2_1$$

لاحظ أن البسط هو الفرق بين المقدرين، ويفترض أن يكون صغيراً تحت  $H_0$  لأنه في هذه الحالة كلا المقدرين متقاربان وبالتالي يفترض أن يتقاربا. بينما المقام هو الفرق بين التباين وهذا يفترض أن يكون كبيراً تحت  $H_0$  لأن تباين مقدر نموذج الخطأ العشوائي يكون أقل منه في نموذج الأثر الثابت. هذا يعني أنه في ظل  $H_0$  فإن البسط صغير والمقام كبير فيأتي الكسر قريباً من الصفر. على العكس في حالة عدم تحقق  $H_0$  نعلم أن البسط يكون كبيراً لأن مقدر نموذج الأثر الثابت فقط يكون متقارباً بينما الآخر غير متقارب، وبالتالي يختلفان عن بعضهما. بينما المقام يكون صغيراً لأن ميزة صغر تباين مقدر نموذج الأثر العشوائي تزول. في المحصلة يستبعد أن تأتي الإحصائية بعيدة عن الصفر في حالة تحقق  $H_0$  لذلك نرفض هذه الأخيرة إذا زادت قيمة الإحصائية عن القيمة الجدولية.

- في حالة كون من المعقول افتراض عدم ارتباط  $\alpha_i$  بالمفسرات، فإن نموذج الأثر العشوائي يعطي إيجابيتين أساسيتين: أن مقدراته متقاربة وأدق (consistent)، أي يعطي خطأ معياري أقل من مقدرات نموذج الأثر الثابت.
- والميزة الثانية لنموذج الأثر العشوائي هي أنه يحتسب المتغيرات الثابتة في الزمن (بما أنه لا يستخدم الفروق عن المتوسط).

عيب نموذج الأثر العشوائي هو عدم تحقق فرضيته في الغالب.

لكن يجب التصحيح من الارتباط الذاتي وعدم التجانس<sup>1</sup> الذي ينشأ من ترك الفروق الفردية في حد الخطأ (خطأ مفردة في فترة غير مستقل عنه في فترة سابقة: إن كانت قيمه أعلى في فترة فهي غالباً أعلى في الفترة الموالية، والعكس...)، فمفردات الخطأ  $u_{it}$ ، بسبب ترك الفروق الفردية ضمنها، ليست (كما نريدها) مستقلة عن بعضها وبنفس التوزيع ومتوسطها الصفر، أحد الفرضيات الأساسية في OLS. برنامج R من خلال plm يقوم بهذا التصحيح تلقائياً لكن يتعين استخدام اختبار هاوسمن للاستقلال بين الفروقات الفردية والمفسرات، إذا جاء الاختبار دالاً فهذا يعني الذهاب إلى طريقة الأثر الثابت وليس الأثر العشوائي.

### ملاحظات

- يمكن اختبار وتصحيح الارتباط الذاتي وعدم التجانس
- يمكن تقدير خطأ معياري مقاوم لكل من الارتباط الذاتي وعدم التجانس
- يمكن تصور نماذج حيث يوجد الأثر الثابت حتى بدون هيكل بيانات البانل التقليدي، مثلاً بيانات التوائم.
- Hausman test في R : phtest

### المقارنة بين FE و RE

#### مثال عن المفاضلة بين طريقة الأثر الثابت والأثر العشوائي<sup>2</sup>.

لدينا نتائج 4 طلبة في 3 سنوات، ونريد اختبار تأثير العلامات (Grade) بعدد ساعات المذاكرة أسبوعياً (StudyTime). نتوقع أن الطلبة مختلفين (هناك فروقات فردية تؤثر على التابع: العلامات)، وهذه الفروقات الفردية نرجح أنها مرتبطة بالمتغيرة المفسرة وهي عدد ساعات المذاكرة، حيث قد يكون الطلبة الأكثر ذكاءً يميلون إلى أن يكونوا مثابرين (عدد ساعات مذاكرة أكبر من غيرهم)، أو العكس لا يحتاجون إلى المذاكرة لساعات طويلة، أي كانت العلاقة طردية أم عكسية. في هذه الحالة نحتاج إلى إدخال هذه الفروقات الفردية في النموذج (من خلال متغيرات وهمية) ولا نهملها، لأنه حينما وجدت متغيرات مهمة لها علاقة بالمتغيرات المفسرة، يكون لدينا تحيز في المعاملات المقدر (تحيز المتغيرات المهملة omitted variables bias).

في الواقع يمكن أن نتصور أن الفروقات الفردية في السنوات أيضاً لها علاقة بالعلامات وبالمفسرة عدد ساعات الدراسة، بحيث أن بعض السنوات أصعب من الأخرى، فهي تؤثر على العلامة وتؤثر على عدد ساعات الدراسة التي يخصصها الطالب في هذه السنة أو تلك.

في هذه الحالة (وجود متغيرات وهمية مرتبطة بالمفسرات) لا يجدر استخدام طريقة الأثر العشوائي، لأن هذه الأخيرة تقترض استقلال الفروقات الفردية عن المفسرات (وتعتبرها عشوائية وبالتالي تتركها مع حد الخطأ أي بدون تقدير أي أنها تهملها) ولا طريقة التقدير البيني للسبب نفسه وهو فرضية الاستقلال عن المفسرات. الطريقة التي تسمح بالارتباط بين الفروق الفردية والمفسرات هي طريقة الفروق الأولى FD أو طريقة الأثر الثابت FE.

إذا أردنا إدراج هذه الفروقات الفردية في الدالة (لكيلا نهملها) (نموذج الفروق الفردية) سنجد أننا نحتاج إلى الكثير من المتغيرات الوهمية (3 متغيرات لـ 4 طلبة متغيرتين لـ 3 اختبارات). هذا يستهلك الكثير من درجات الحرية وبالتالي يضخم الخطأ المعياري للمقدر (الخطأ المعياري للمقدر يتأثر بحجم العينة مقارنة مع عدد المتغيرات المفسرة). طريقة الأثر الثابت

<sup>1</sup>Non-spherical Error variance

<sup>2</sup>BurkeyAcademy, 2014, (02-04-2010), <https://www.youtube.com/channel/UCVjJYEZwPr-Q1yqyQsELC3g>

تحل هذه المشكلة فهي تسحبها من الدالة لأنها تستخدم الفرق عن المتوسط في كل فترة وبما أن المتغيرات الوهمية ثنائية فهي في كل فترة إما أن تأخذ 0 أو 1 أي نفس القيمة خلال الفترة وبالتالي فالفرق تعطي أصفارا (0-0=0, 1-1=0)

Student	Year	StudyTime	Grade	Jamel	Sali	Mabrouk	Year2	Year3
Antar	1	8	66.5	0	0	0		
Antar	2	5	50.4	0	0	0		
Antar	3	9	69.0	0	0	0		
Jamel	1	4	54.7	1	0	0		
Jamel	2	6	60.3	1	0	0		
Jamel	3	2	38.3	1	0	0		
Sali	1	11	86.1	0	1	0		
Sali	2	3	45.3	0	1	0		
Sali	3	7	64.3	0	1	0		
Mabrouk	1	2	48.9	0	0	1		
Mabrouk	2	1	39.1	0	0	1		
Mabrouk	3	2.5	46.8	0	0	1		

لكن إذا افترضنا أن الفروق الفردية بين الطلبة ليست مرتبطة بالمفسرات المدرجة، عندها يمكن استخدام طريقة الأثر العشوائي، أي يمكنك إهمال المتغيرات الوهمية التي تمثل الفروقات الفردية بدون أن يؤدي ذلك لتحيز المقدرات. طريقة الأثر العشوائي إذن تترك المتغيرات الوهمية الثنائية التي تلتقط الفروقات الفردية فهي يعتبرها ضمن الخطأ العشوائي وبالتالي فهي لن تظهر في الدالة ولا نحتاج لتقديرها.

#### خلاصة عامة

- يسمح استخدام البيانات المقطعية المجمعَة باختبار وجود تغيرات هيكلية
- طريقة بسيطة في البيانات المقطعية المجمعَة لتقييم أثر الأحداث والسياسات الاقتصادية هي فرق الفروق
- بيانات البنابل تسمح بمعالجة مشكلة التحيز الناجم عن المتغيرات غير المشاهدة
- طريقة الأثر الثابت تعطي مقدرات متقاربة ولا يتأثر ذلك بارتباط الفروقات الفردية غير المشاهدة بالمفسرات
- طريقة الأثر العشوائي هي أكثر فعالية في حالة غياب هذا الارتباط بين الفروقات الفردية والمفسرات، لكن هذا الافتراض قل ما يتحقق.