

EViewsによる パネルデータの分析

LightStone Corp

EViewsパネル分析

■ 講座の目的

- パネルデータをEViewsで構造化でき、固定効果およびランダム効果モデルの推定ができるようになること。また、各種検定手法を用いて適切なモデルを選択することができるようになる事を目的とします。

■ 構成

- EViewsにおけるパネルワークファイルの作成
 - ヘドニック価格(1時点)の推定と固定効果モデル
 - ランダム効果モデルの推定とハウスマン検定
 - 一人当たりGDPと人口増加率の関係(時系列)
 - ダイナミックパネル
- フォルダevIEWS06のファイルを利用します。

クロスセクションID

- パネルワークファイルでは必ず個体を識別するIDが必要です。

操作1: EXCELサンプルファイルbank.xlsをEXCELで開いて内容を確認します。

bank: 個体を識別するIDです。これをクロスセクションIDと呼びます。

year: 時間を識別するシリーズです。

操作2: このファイルを閉じます。そして、EViewsで改めて開きます。

単純にデータが12個ならんだものとして、EViewsは認識します。

	A	B	C	D
	bank	year	employee	capital
2	aomori	1999	235	12
3	aomori	2000	211	12
4	aomori	2001	200	11
5	iwate	1999	212	13
6	iwate	2000	201	15
7	iwate	2001	189	15
8	akita	1999	199	10
9	akita	2000	167	10
10	akita	2001	151	10
11	miyagi	1999	320	23
12	miyagi	2000	301	21
13	miyagi	2001	298	21

パネルワークファイルのスタック形式

ワークファイルの構造

- パネルワークファイルの自動認識(EViews7)。

操作1: Rangeの表示がDated Panelになっています。

Range: 1999 2001 × 4 -- 12 obs
Sample: 1999 2001 -- 12 obs

操作2: Rangeの表示をダブルクリックしてクロスセクションIDと日付IDを確認します。

操作3: ワークファイルを「bank」という名前で保存して閉じます。

The screenshot shows the 'Workfile Structure' dialog box in EViews 7. The 'Workfile structure type' is set to 'Dated Panel'. Under 'Panel identifier series', the 'Cross section ID series' is 'bank' and the 'Date series' is 'dateid'. Under 'Observation inclusion/creation', the 'Frequency' is 'Annual', the 'Start date' is '@first', and the 'End date' is '@last'. There are checkboxes for 'Balance between starts & ends' (checked), 'Balance starts', 'Balance ends', and 'Insert obs to remove date gaps so date follows regular frequency' (unchecked). 'OK' and 'Cancel' buttons are at the bottom.

アンスタックデータ

- アンスタックデータを構造化してみましょう

操作1: unstack_bank1を開いて内容を確認します。

操作2: ファイルを直接EViewsで開きます。

Range: 1 12 Dim(3,4) -- 12 obs
Sample: 1 12 -- 12 obs

操作3: パネルワークファイルとして構造化してください。

操作4: employeeのシリーズオブジェクトを開いて右図のように構造化されていることを確認しましょう。ファイルを閉じます。

	A	B	C	D
1	bank	year	employee	capital
2	aomori	1999	235	12
3	iwate	1999	212	13
4	akita	1999	199	10
5	miyagi	1999	320	23
6	aomori	2000	211	12
7	iwate	2000	201	15
8	akita	2000	167	10
9	miyagi	2000	301	21
10	aomori	2001	200	11
11	iwate	2001	189	15
12	akita	2001	151	10
13	miyagi	2001	298	21

Yearがきちんと揃っています。

		Last u
		Imported from 'C
akita - 99	199	
akita - 00	167	
akita - 01	151	
aomori - 99	235	
aomori - 00	211	
aomori - 01	200	
iwate - 99	212	
iwate - 00	201	
iwate - 01	189	
miyagi - 99	320	
miyagi - 00	301	
miyagi - 01	298	

欠損値

- EXCELファイルmissing1とmissing2を使って欠損値の取り扱いを確認します。

bank	year	employee	capital
aomori	1999	235	12
iwate	1999	212	13
akita	1999	199	10
miyagi	1999	320	23
tokyo	1999	2262	123
aomori	2000	211	12
iwate	2000	201	15
akita	2000	167	10
miyagi	2000	301	21
aomori	2001	200	11
iwate	2001	189	15
akita	2001	151	10
miyagi	2001	298	21

Missing1: 必要なセルにデータを入れました。

bank	year	employee	capital
aomori	1999	235	12
iwate	1999	212	13
akita	1999	199	10
miyagi	1999	320	23
tokyo	1999	2262	123
aomori	2000	211	12
iwate	2000	201	15
akita	2000	167	10
miyagi	2000	301	21
tokyo	2000		
aomori	2001	200	11
iwate	2001	189	15
akita	2001	151	10
miyagi	2001	298	21
tokyo	2001		

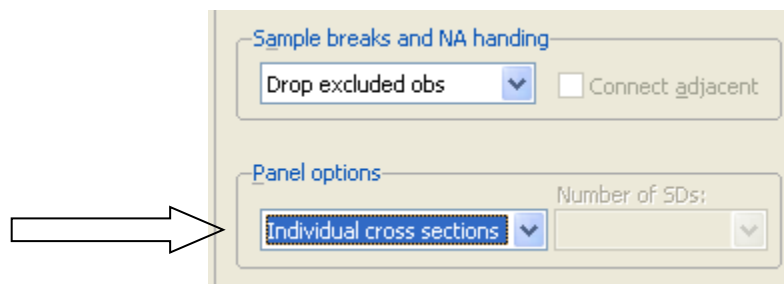
Missing2: 余分なセルも用意し空白にしておきます。

操作: 2つのファイルを別ページで開いて、パネルワークファイルに構造化してください。

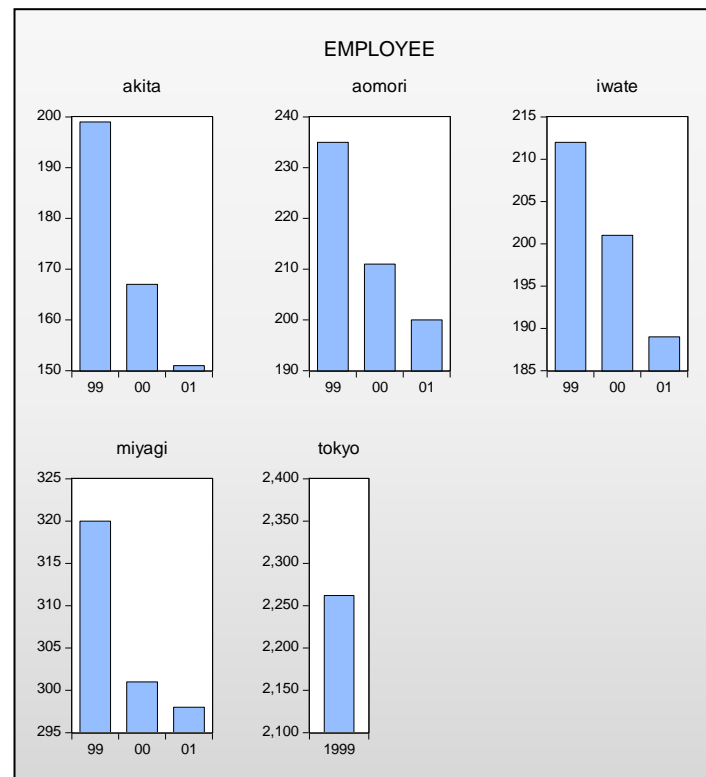
グラフの作成

- パネルワークファイルにはクロスセクションの利用方法により、色々なグラフを作成できます。

操作: シリーズemployeeで色々なグラフを作ってみましょう。



オプションを一通り選択して、
グラフの種類を確認しましょう。



失敗例

- データをうまくパネルワークファイルに変換できない場合。

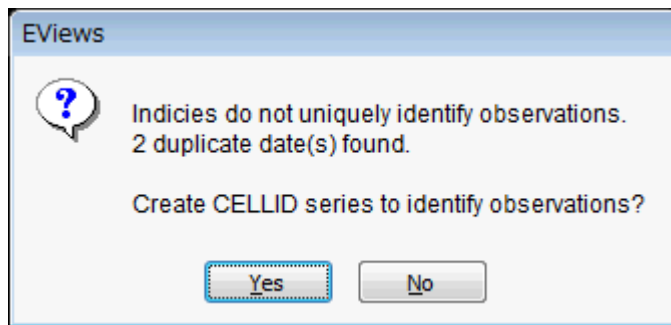
個体を識別するためのIDが重複している事があります。この場合は構造化できません。

操作: Excelデータbankng.xlsを開いて、データが重複している事を確認します。

パネルデータは一般的に大容量ですので、このように目視で誤りを見つけるのは困難です。

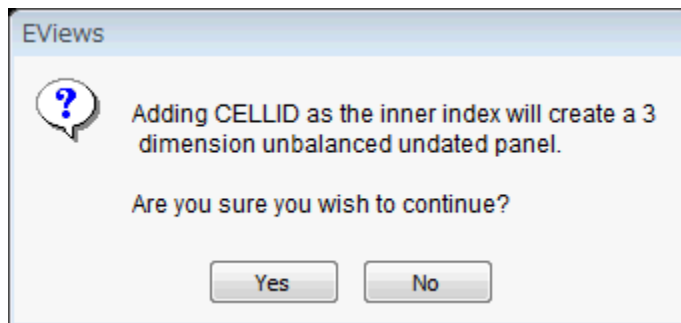
失敗例

操作1: 前述の方法でbankng.xlsをEViewsで開いてパネルワークファイルとして構造化しますと次のメッセージを表示します。



重複したインデックスがある、とのメッセージを表示します。

操作2: そのままOKボタンをクリックします。



セルIDを利用してアンバランスな日付未対応のパネルデータを作成するというメッセージを表示します。

操作3: 同じくOKボタンをクリックします。

重複したデータ

Range: 1 14 Dim(4,3,2) -- 14 obs
Sample: 1 14 -- 14 obs

パネルデータとして認識できていません。14個のデータのある通常のワークファイルを作成しました。

操作1: bankとyearのグループオブジェクトgroup01を作成します。

操作2: group01でView/N-way tabulationと操作します。Output項目はCountのみチェックを残してOKボタンをクリックします。

		YEAR			
Count		1999	2000	2001	Total
BANK	akita	1	1	1	3
	aomori	2	2	1	5
	iwate	1	1	1	3
	miyagi	1	1	1	3
	Total	5	5	4	14

Bank:aomori, Year:1999と2000に重複が認められます。

重複したデータ

*重複したデータを削除する場合は、Workfile StructureをUndatedに変更した上で、削除作業を行います。

*シリーズオブジェクト、グループオブジェクトのどちらからでも削除できます。

定式化の誤り

■ 変数の過不足とその影響

- 2つの回帰モデルを用いて基礎知識を確認しましょう。

単回帰モデル $Y_i = \alpha + \beta X_{2i} + u_i$ (1)

$$\hat{\beta} = \frac{\sum (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}_2)}{\sum (X_{2i} - \bar{X}_2)^2} = \frac{\sum (X_{2i} - \bar{X}_2)Y_i}{\sum (X_{2i} - \bar{X}_2)^2} \quad (2)$$

多重回帰モデル $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ (3)

$$\hat{\beta}_2 = \frac{\sum \hat{v}_i Y_i}{\sum \hat{v}_i^2} \quad (4)$$

$$X_{2i} = \gamma_1 + \gamma_2 X_{3i} + v_i$$

過少定式化の誤り

- 変数を少なく見積もってしまった場合

$$\hat{\beta} = \frac{\sum (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}_2)}{\sum (X_{2i} - \bar{X}_2)^2} = \frac{\sum (X_{2i} - \bar{X}_2)Y_i}{\sum (X_{2i} - \bar{X}_2)^2} \quad (2)$$

2式に3式を代入すると不偏推定量でない事がわかる

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (3)$$

$$\hat{\beta} = \beta_2 + \frac{\sum (X_{2i} - \bar{X}_2)(\beta_3 X_{3i} + u_i)}{\sum (X_{2i} - \bar{X}_2)^2}$$

$$E(\hat{\beta}) = \beta_2 + \frac{\beta_3 \sum (X_{2i} - \bar{X}_2)X_{3i}}{\sum (X_{2i} - \bar{X}_2)^2} \neq \beta_2$$

過少定式化の誤り

- 一致性を持つためには次の条件を満たす必要がある

$$\lim_{n \rightarrow \infty} \left\{ \frac{\beta_3 \sum (X_{2i} - \bar{X}_2) X_{3i}}{\sum (X_{2i} - \bar{X}_2)^2} \right\} = 0$$

一般的にこのような事は成立しないので、一致性は無い

結論：過少定式化の場合、不偏性も一致性もない

過剰定式化の誤り

■ 過剰定式化(余計に説明変数を使ったとき)

単回帰モデル $Y_i = \alpha + \beta X_{2i} + u_i$ (真のモデル)

多重回帰モデル $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

過少定式化と同じ考え方で推定値を求めると、

$$\hat{\beta}_2 = \beta + \frac{\sum \hat{v}_i u_i}{\sum \hat{v}_i^2}$$

期待値は $E(\hat{\beta}_2) = \beta \leftarrow$ 不偏推定量

分散は $\sigma_{\hat{\beta}_2}^2 = \frac{\sigma^2}{\sum \hat{v}_i^2} > \frac{\sigma^2}{\sum (X_{2i} - \bar{X}_2)^2}$ 有効性はない

定式化の誤り

- 過少定式化(説明変数を取り落としている時)

不偏性も一致性もありません

- 過剰定式化(説明変数を余計に利用している時)

有効性はありませんが、不偏性があります。

Point: 変数の選択により、決定係数が左右されるという事に加えて、推定量の性質は大きく影響される。少ないよりは、多い方が良い。

脱落変数の検定

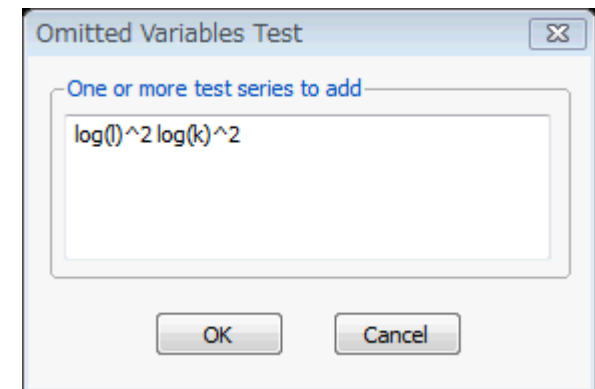
■ EViewsの脱落変数の検定の操作方法を学びましょう

操作1:ワークファイルcoef_test.wf1を開きます。コブ-ダグラス型の生産関数を推定します。

操作2:メインメニューからQuick/Estimate Equationと操作して次のモデルを推定し、eq01という名前を付けます。

$$\log(q) = c + \log(l) + \log(k)$$

操作3:この推定式に $\log(l)^2$ と $\log(k)^2$ を追加すべきか、検定を行います。eq01でView/Coefficient Diagnostics/Omitted Variables-Likelihood Ratio...と操作します。スペース区切りで2つの変数名を入力します。



尤度比検定

■ 脱落変数の検定結果

帰無仮説は「追加した変数のパラメータは共にゼロである(ムダである)」です。有意水準5%では、帰無仮説は棄却できません。

$$LR = -2(l_r - l_u)$$

尤度比検定の自由度
は制約の数です。

Omitted Variables Test			
Equation: EQ01			
Specification: LOG(Q) C LOG(L) LOG(K)			
Omitted Variables: LOG(L)^2 LOG(K)^2			
	Value	df	Probability
F-statistic	2.490982	(2, 20)	0.1082
Likelihood ratio	5.560546	2	0.0620
F-test summary:			
	Sum of Sq.	df	Mean Squares
Test SSR	0.008310	2	0.004155
Restricted SSR	0.041669	22	0.001894
Unrestricted SSR	0.033359	20	0.001668
Unrestricted SSR	0.033359	20	0.001668
LR test summary:			
	Value	df	
Restricted LogL	44.48746	22	
Unrestricted LogL	47.26774	20	

パネルデータのメリット

■ 異質性

個人、企業、地域(都道府県)の異質性を捕えることができます。
異質性を考慮しない時系列やクロスセクションの分析は、バイアスのかかった結果をもたらすことが考えられます。

■ Baltagi and Levin(1992)より

合衆国における1963-88年のタバコの需用関数の推定

✓タバコの消費を被説明関数として、ラグ付き消費、価格、所得を説明変数としてタバコの需要関数を推定し、禁煙キャンペーンの効果を実証分析しました。

✓タバコの消費に影響を与える変数は、これら以外にも考えられます。
例えば、モルモン教徒は喫煙は禁止されています(ユタ州)。

固定効果モデル

2変数による回帰モデル(プーリング推計)

$$y_t = \alpha + \beta x_t + u_t$$

説明変数Xに複数の種類(個体:例えば銀行など)がある。

$$y_t = \alpha + \alpha_i + \beta x_{it} + u_{it}$$

ホワイトノイズ

時間には無関係な、個体(i)固有の効果 α_i がある(固定効果)。

固定効果モデルを推定したら、 α_i が冗長な変数になっていないか、Redundant Variable Testで検定します。

固定効果モデルの推定

- Harrison and Rubinfeld(1978) のヘドニック価格に関する分析を行います(harrison_panel.wf1)。

分析の内容

住宅(持ち家)価格の中央値(MV:対数値)を被説明変数として、説明変数には住宅価格に影響を与える次のような変数を利用します。ボストン近郊の92の町から合計506の地域でデータを集めました。

住宅のヘドニック価格 ←

狙い: 町の異質性を固定効果として捉える

1時点のデータなので、定常性を考える必要はありません。

CRIM: 犯罪率

CHAS: Charles Riverに隣接することを示すダミー変数

NOX: 大気汚染

RM: 平均部屋数

AGE: older unitsの比率

DIS: 職業安定所からの距離

B: アフリカ系アメリカ人の比率

LSTAT: 低所得者の比率

クロスセクション

ワークファイルを構造化せずにそのまま重回帰モデルを推定します。

```
Range: 1 506 -- 506 obs
Sample: 1 506 -- 506 obs
```

操作: mvを被説明変数として次の重回帰式を推定します。
Equationオブジェクトの名前は「eq01」とします。

```
mv c crim chas nox rm age dis b lstat
```

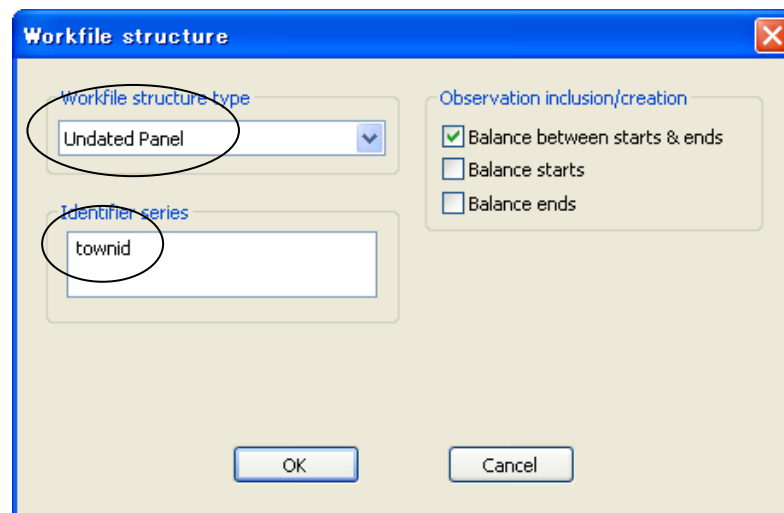
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.978637	0.110983	80.90115	0.0000
CRIM	-1.213214	0.120733	-10.04874	0.0000
CHAS	1.235372	0.350433	3.525272	0.0005
NOX	-0.471710	0.108320	-4.354765	0.0000
RM	0.849802	0.135772	6.259045	0.0000
AGE	-0.171440	0.553656	-0.309651	0.7570
DIS	-1.649795	0.320473	-5.147995	0.0000
B	0.387182	0.108461	3.569782	0.0004
LSTAT	-3.939697	0.262811	-14.99059	0.0000

プーリング推定

構造化

地域を識別するtownidを使って、クロスセクションのデータとして分析します。

操作: townidを利用してパネルワークファイル(クロスセクションのみ)として構造化します。セルIDの作成を確認するメッセージを2つ表示しますが、そのまま、OKボタンをクリックして構造化します。



構造化

92のtownidでデータを識別しました。1つの町で最大30個の地域をサンプリングしています。

Range: 1 506 Dim(92,30) -- 506 obs
Sample: 1 506 -- 506 obs

操作: シリーズtownidでview/
One-Way tabulationを選択して
確認します。
29番目の地域で、30地域の
データを調査しています。

20	4	0.79	88	17.39
21	4	0.79	92	18.18
22	3	0.59	95	18.77
23	5	0.99	100	19.76
24	11	2.17	111	21.94
25	9	1.78	120	23.72
26	7	1.38	127	25.10
27	15	2.96	142	28.06
28	30	5.93	172	33.99
29	7	1.38	179	35.38
30	8	1.58	187	36.96
31	6	1.19	193	38.14
32	2	0.40	195	38.54
33	1	0.20	196	38.74
34	3	0.59	199	39.33
35	2	0.40	201	39.72
36	2	0.40	203	40.12
37	2	0.40	205	40.51
38	2	0.40	205	40.51
39	11	2.17	216	42.69

モデル推定

操作1: Quick/Estimate Equationと操作して次のように入力し、推定します。
推定式のオブジェクト名をeq02とします。変数はeq01と同じです。

mv c crim chas nox rm age dis b lstat

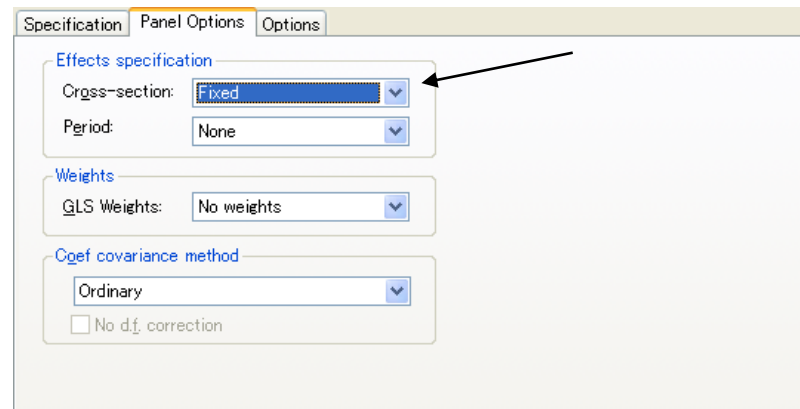
パネルデータとして設定しても「効果」を利用しない場合、パラメータと統計量はeq01と同じになります。

	Coefficient	Std. Error	t-Statistic	Prob.
C	8.978637	0.110983	80.90115	0.0000
CRIM	-1.213214	0.120733	-10.04874	0.0000
CHAS	1.235372	0.350433	3.525272	0.0005
NOX	-0.471710	0.108320	-4.354765	0.0000
RM	0.849802	0.135772	6.259045	0.0000
AGE	-0.171440	0.553656	-0.309651	0.7570
DIS	-1.649795	0.320473	-5.147995	0.0000
B	0.387182	0.108461	3.569782	0.0004
LSTAT	-3.939697	0.262811	-14.99059	0.0000

$$Y_i = a + bX_i + u_i$$

固定効果モデルの推定

操作: eq02のEstimateボタンをクリックして、Panel Optionsのタブを表示します。Cross-sectionのタブでFixedを選択して、OKボタンをクリックします。



AGEは有意になりましたが、今度はCHASとDISが有意でないという結果になりました。

	Coefficient	Std. Error	t-Statistic	Prob.
C	8.993272	0.134738	66.74632	0.0000
CRIM	-0.625400	0.104012	-6.012746	0.0000
CHAS	-0.452414	0.298531	-1.515467	0.1304
NOX	-0.558938	0.135011	-4.139949	0.0000
RM	0.927201	0.122470	7.570833	0.0000
AGE	-1.406955	0.486034	-2.894767	0.0040
DIS	0.801437	0.711727	1.126045	0.2608
B	0.663405	0.103222	6.426958	0.0000
LSTAT	-2.453027	0.255633	-9.595892	0.0000

$$Y_{it} = a + bX_{it} + v_i + \epsilon_{it}$$

固定効果モデルの判定

- 固定効果「なし」のモデルと「有り」のモデルの選択

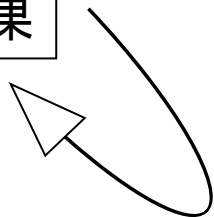
操作: eq02でView: Fixed/Random Effects – Testing/Redundant Fixed Effects – Likelihood Ratioと操作します。

Redundant Fixed Effects Tests			
Equation: EQ01			
Test cross-section fixed effects			
Effects Test	Statistic	d.f.	Prob.
Cross-section F	7.763171	(91,406)	0.0000
Cross-section Chi-square	510.030613	91	0.0000

帰無仮説は「固定効果は冗長である」です。p値がゼロとなっていますので、棄却されます。つまり、「固定効果」は有効だということが分かります。

固定効果

効果なし



検定の結果について

- プーリング推定と固定効果モデルの結果を比較してみましょう。

CRIM...どちらも有意

CHAS...川からの距離は実は関係ありません。

NOX...どちらも有意

RM...どちらも有意

AGE...古いunitの比率は最初、関係ないようでしたが、実は有意です。

DIS...逆に、雇用センターからの距離は実は有意ではありません。

B...どちらも有意

LSTAT...どちらも有意

固定効果モデルで有意ではないCHASとDISは、式から取り去ったほうがいいのでしょうか？

	Coefficient	Std. Error	t-Statistic	Prob.
C	8.978637	0.110983	80.90115	0.0000
CRIM	-1.213214	0.120733	-10.04874	0.0000
CHAS	1.235372	0.350433	3.525272	0.0005
NOX	-0.471710	0.108320	-4.354765	0.0000
RM	0.849802	0.135772	6.259045	0.0000
AGE	-0.171440	0.553656	-0.309651	0.7570
DIS	-1.649795	0.320473	-5.147995	0.0000
B	0.387182	0.108461	3.569782	0.0004
LSTAT	-3.939697	0.262811	-14.99059	0.0000

プーリング推計

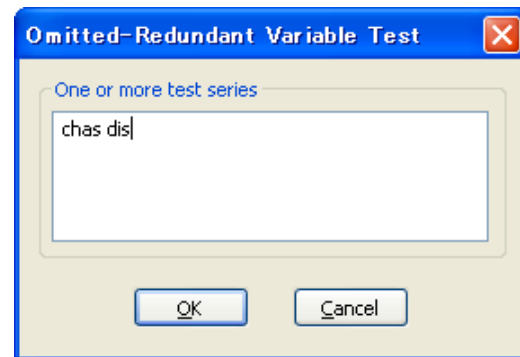
	Coefficient	Std. Error	t-Statistic	Prob.
C	8.993272	0.134738	66.74632	0.0000
CRIM	-0.625400	0.104012	-6.012746	0.0000
CHAS	-0.452414	0.298531	-1.515467	0.1304
NOX	-0.558938	0.135011	-4.139949	0.0000
RM	0.927201	0.122470	7.570833	0.0000
AGE	-1.406955	0.486034	-2.894767	0.0040
DIS	0.801437	0.711727	1.126045	0.2608
B	0.663405	0.103222	6.426958	0.0000
LSTAT	-2.453027	0.255633	-9.595892	0.0000

固定効果モデル

変数の削除

■ 冗長な変数の削除を検定する

操作1: eq02でView/Coefficient Diagnostics/Redundant Variable - Likelihood Ratio...と操作します。ダイアログにスペース区切りでchas dis と入力します。



Equation: EQ02
Specification: MV C CRIM CHAS NOX RM AGE DIS B LSTAT
Redundant Variables: CHAS DIS are jointly insignificant

	Value	df	Probability
F-statistic	1.812605	(2, 406)	0.1645
Likelihood ratio	4.498067	2	0.1055

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.061501	2	0.030750
Restricted SSR	6.949184	408	0.017032
Unrestricted SSR	6.887683	406	0.016965

帰無仮説は「これら2つの変数は元の式で冗長である」です。2つの検定統計量とも、有意水準10%としても、帰無仮説を棄却できません。

操作2: eq02をオブジェクトコピーして「eq02fix」を作成し、2つの変数を削除します。

固定効果モデル

住宅のヘドニック価格

□地域の異質性を考慮して考えるべき。(固定効果モデル)

□チャールズ川からの距離は価格には関係ない。

□雇用センターからの距離も関係ない。

□Older unitは過小評価されていた。この変数はモデルで利用すべきである。

	Coefficient	Std. Error	t-Statistic	Prob.
C	9.116166	0.088220	103.3342	0.0000
CRIM	-0.641630	0.103426	-6.203768	0.0000
NOX	-0.621754	0.124992	-4.974333	0.0000
RM	0.921504	0.120925	7.620459	0.0000
AGE	-1.540902	0.480900	-3.204204	0.0015
B	0.650928	0.103158	6.310040	0.0000
LSTAT	-2.479254	0.255701	-9.695909	0.0000

Eq02FIXの推定結果画面

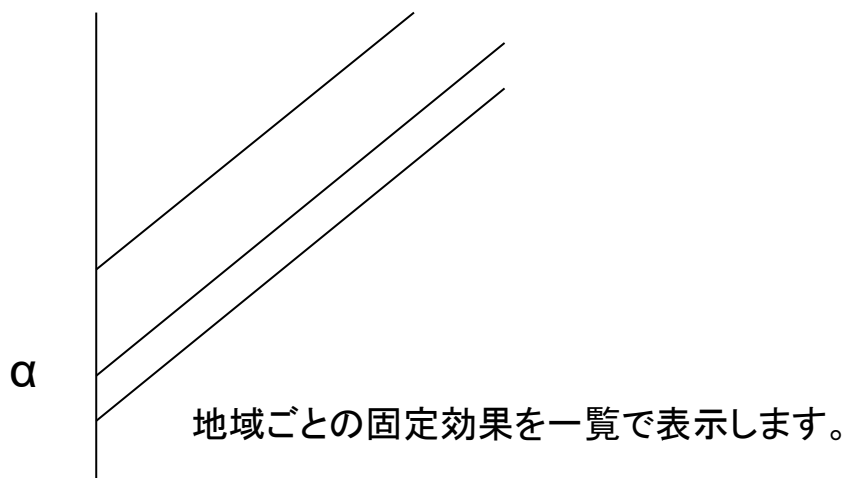
ヘドニック価格を決めるにあたり、計測可能な数値データだけでなく、観測できない地域ごとの「異質性」を固定効果として取り入れて、モデルを考える方が妥当である。

固定効果を見る

- 固定効果モデルにおける各個体のごとの効果を調べる。

$$Y_{it} = a + bX_{it} + \underbrace{v_i}_{\text{固定効果}} + \epsilon_{it}$$

操作：推定が完了したeq02fixのViews
ボタンでFixed/Random Effect : Cross-
section Effectと操作します。



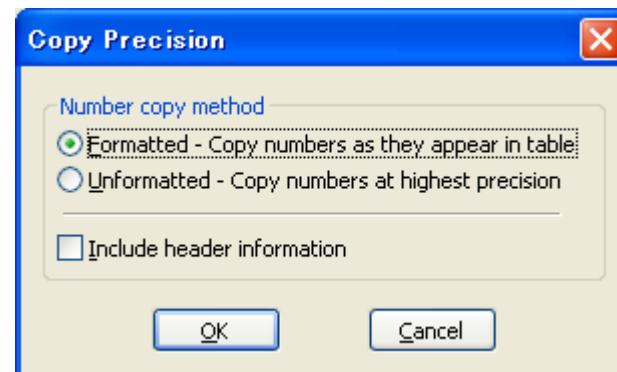
Cross-section Fixed Effects			
	TOWNID	Effect	
1	1.0000	-0.150580	
2	2.0000	-0.033450	
3	3.0000	0.019024	
4	4.0000	0.053053	
5	5.0000	-0.134583	
6	6.0000	-0.076052	
7	7.0000	-0.136639	
8	8.0000	-0.129450	
9	9.0000	-0.149470	
10	10.000	-0.142759	
11	11.000	-0.001082	
12	12.000	-0.183084	
13	13.000	-0.072886	
14	14.000	-0.155330	
15	15.000	0.141963	
16	16.000	-0.268253	
17	17.000	-0.204677	

固定効果の大きさ

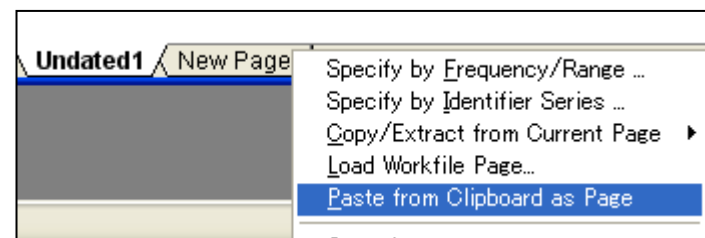
■ 観測できない固定効果(異質性)の一番大きな町は？

固定効果の表を他のワークファイルページに貼り付けて、sortコマンドで並べ替えて調べます。

操作1: 列を選択してコピーします。右のダイアログではそのままOKボタンをクリックします。



操作2: ワークファイルのタブから「Paste from Clipboard Page」を選択して、新しいページに貼り付けます。

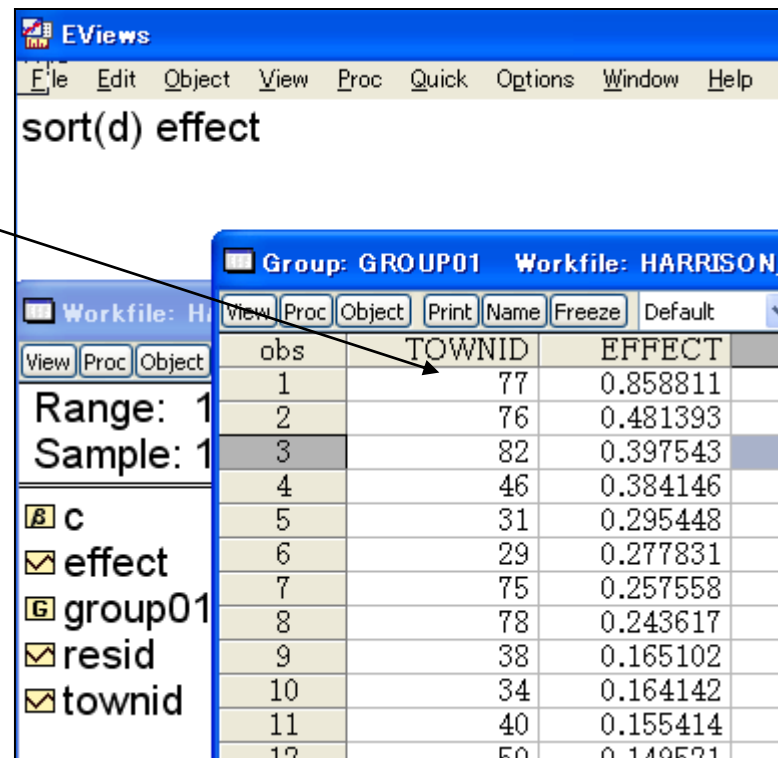


固定効果の考察

操作1: コマンドウィンドウで「sort(d) effect」として、効果の大きさをキーにして、降順で並べ替えます。

townid 77番の町の固定効果が飛び抜けて大きいことが分かりました。ただし、固定効果モデルの
変量効果に**一**致性はありません

操作2: effectとtownidのグループ名をgroup01として、ワークファイルを上書き保存します。



obs	TOWNID	EFFECT
1	77	0.858811
2	76	0.481393
3	82	0.397543
4	46	0.384146
5	31	0.295448
6	29	0.277831
7	75	0.257558
8	78	0.243617
9	38	0.165102
10	34	0.164142
11	40	0.155414
12	50	0.149521

まとめ

■ ヘドニック価格

- townidを使ってパネルワークファイルに構造化
- プーリング推計: AGEは有意ではない
- 固定効果モデル: CHASとDISが有意ではない
- Redundant Testの結果、固定効果モデルを採用。
AGEは有意であり、CHASとDISは有意ないという結論を得ました。
- 異質性を示す固定効果の最も大きな町はTOWNID 77番。

練習

■ 賃金と人種の関係を分析する

□ サンプルデータ: lnwage_panel.wf1

米国のある年(1年)の賃金のデータを次のモデルで推定します。

被説明変数: 賃金(lnwage)

説明変数: ed(学歴)、 age(年齢)、 asian(アジア系を示すダミー変数)、
定数項

クロスセクションID: gmstcen

分析の狙い

アジア系であることが賃金にどのように影響しているかを調べます。

分析結果

1.効果なし:アジア系であることが賃金に正の影響を与えています。

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.080370	0.015618	-5.146007	0.0000
ED	0.115398	0.001035	111.5470	0.0000
AGE	0.025093	0.000232	108.2062	0.0000
ASIAN	0.025921	0.014093	1.839287	0.0659

2.地域ごとの特徴を固定効果として推定すると、アジア系であることが負の影響を与えることが分かります。

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.072181	0.015613	-4.623249	0.0000
ED	0.115307	0.001035	111.3963	0.0000
AGE	0.024961	0.000231	108.0852	0.0000
ASIAN	-0.017936	0.014771	-1.214336	0.2246

しかし、有意では
なくなりました！

3.賃金は人種には関係なく、教育水準と年齢、そして地域の特徴によって考える方が妥当なようです。

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.071729	0.015608	-4.595630	0.0000
ED	0.115216	0.001032	111.5993	0.0000
AGE	0.024962	0.000231	108.0882	0.0000

固定効果について

- 企業の投資額(inv)、株価(v)、資本金(k)、10社、20年分のデータについて次のモデルを推定し、固定効果について具体的に考えます。
 - データ: grunfeld.wf1

$$INV = C + V + K$$

操作1: ファイルを開き、企業のIDはシリーズオブジェクト「id」、時間は「t」としてワークファイルを構造化します。

操作2: 固定効果モデルとして上記のモデルを推定します。オブジェクト名は「eq01」とします。固定効果モデルを採用すべきか、検定によって調べてください。

Range: 1 20 x 10 -- 200 obs
Sample: 1 20 -- 200 obs

☐ c
☒ dateid
☒ i

固定効果

■ EQ01の定数項と固定効果

EQ01の定数項は「37.84862」で、固定効果は次のようになっています。

操作1: EQ01でView/Fixed/Random Effects/Cross Section Effectsと操作します。
操作2: Excelを起動し、右図のデータを貼り付けます。各行の値に定数項を足します。

	A	B	C	D	E
1	ID	Effect	C	total	d
2	1	304.9696	37.84862	342.8182	342.8183
3	2	253.5165	37.84862	291.3651	291.3651
4	3	-92.141	37.84862	-54.2924	-54.2924
5	4	0.473526	37.84862	38.32215	38.32214
6	5	-159.761	37.84862	-121.912	-121.912
7	6	-22.9511	37.84862	14.89755	14.89754
8	7	-108.796	37.84862	-70.9471	-70.947
9	8	-29.3383	37.84862	8.51034	8.510337
10	9	-108.741	37.84862	-70.892	-70.892
11	10	-37.2325	37.84862	0.61617	0.616171
12					37.84863
13					

	ID	Effect
1	1	304.9696
2	2	253.5165
3	3	-92.14101
4	4	0.473526
5	5	-159.7606
6	6	-22.95107
7	7	-108.7957
8	8	-29.33828
9	9	-108.7406
10	10	-37.23245

この値と同じものをダミー変数を利用したプーリング推計のモデルで作成します。

ダミー変数

操作1: EQ01をオブジェクトコピーしてEQ02を作成します。
操作2: 10社分のダミー変数を作成します。コマンドウィンドウに次のように入力します。

series d1=(id=1)

操作3: d1の中身を確認します。IDが1の部分だけ1が入っていることを確認したら、これをi=10まで繰り返し、10個のダミー変数を作成します。

Range: 1 20 x 10 -- 200 obs	
Sample: 1 20 -- 200 obs	
<input type="checkbox"/> c	<input checked="" type="checkbox"/> t
<input checked="" type="checkbox"/> d1	<input checked="" type="checkbox"/> v
<input checked="" type="checkbox"/> d10	
<input checked="" type="checkbox"/> d2	
<input checked="" type="checkbox"/> d3	
<input checked="" type="checkbox"/> d4	
<input checked="" type="checkbox"/> d5	
<input checked="" type="checkbox"/> d6	
<input checked="" type="checkbox"/> d7	
<input checked="" type="checkbox"/> d8	
<input checked="" type="checkbox"/> d9	
<input checked="" type="checkbox"/> dateid	
<input type="checkbox"/> eq01	
<input type="checkbox"/> eq02	

ダミー変数によるプーリング推計

操作1: EQ02で次のようにダミー変数を利用します。定数項の「C」は使いません。

Equation specification

Dependent variable followed by list of regressors and PDL terms, OR an explicit equation like

```
inv d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 v k
```

操作2: Panel Optionsタブでクロスセクションの効果をNoneに戻して推定します。

操作3: 右図の結果がEXCELファイルにコピーしたものと一致していることを確認してください。

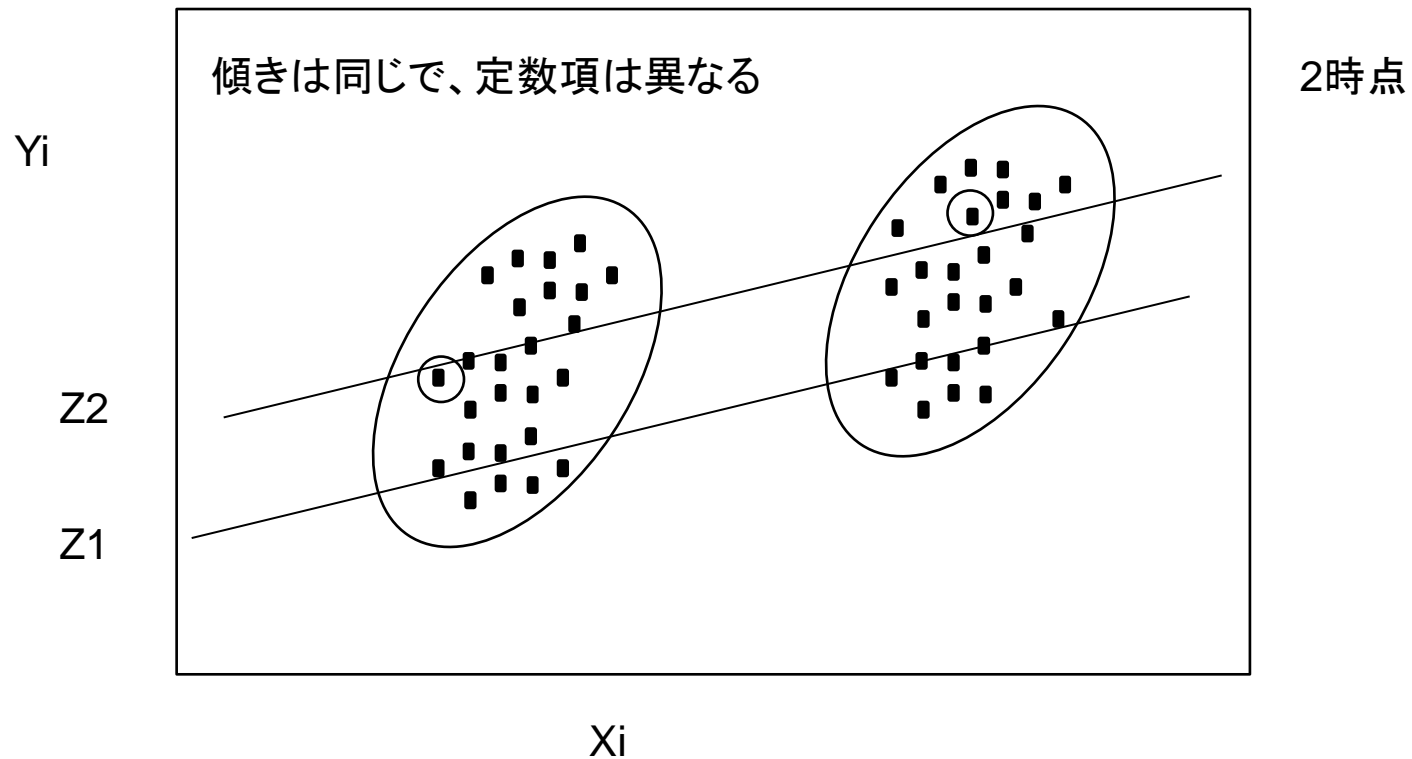
固定効果

- プーリング推計した定数項を、すべての個体に共通する部分と、個体ごとのダミー変数(定数項)に分けることができました。
- 固定効果はダミー変数に対応するパラメータであることが分かりました。
- 個体ごとに有意に、ダミー変数の値が異なる場合は、プーリング推計ではなく、固定効果モデルを用います。

操作: EQ01をオブジェクトコピーしてPeriod Effect(Fix)のみのモデルEQ03を作成し、Redundant Testにより固定効果モデル、プーリング推計のどちらを採用すべきか、調べてください。

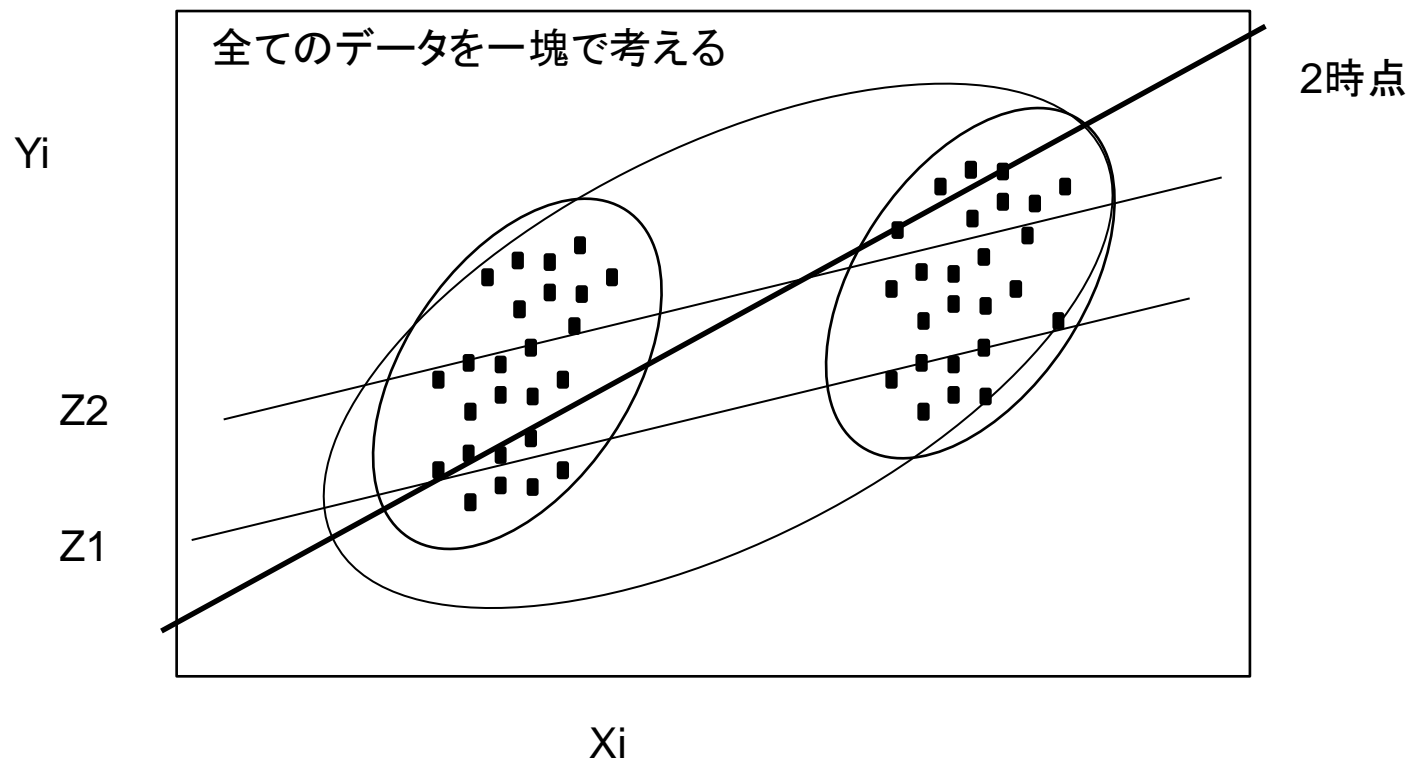
パネルデータ

■ Y_i と X_i の関係



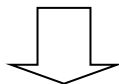
プーリング推定

■ 全体の特徴を見てしまう



回帰モデル

$$Y_i = a + bX_i + u_i$$



$$Y_{it} = \textcircled{a} + bX_{it} + \textcircled{v_i} + \epsilon_{it}$$

(1)

v_i 個体固有の残差: 時間を通じて一定

ϵ_{it} 平均ゼロ、自己相関なし、 X_{it} との相関なし

Within推定量とBetween推定量

式(1)より、時間軸での平均を考えると(Between推定)、

$$\bar{Y}_i = a + b\bar{X}_i + v_i + \bar{\epsilon}_i \quad (2)$$

式(1)から式(2)を引いて(Within推定)、

$$\begin{array}{rcl} Y_{it} & = & a + bX_{it} + v_i + \epsilon_{it} \\ \text{---)} & & \bar{Y}_i = a + b\bar{X}_i + v_i + \bar{\epsilon}_i \\ \hline (Y_{it} - \bar{Y}_i) & = & (X_{it} - \bar{X}_i)b + (\epsilon_{it} - \bar{\epsilon}_i) \end{array} \quad (3)$$

固定効果推定

- Z_i を個体の確定的な要因と考える。

$$Y_{it} = a + bX_{it} + Z_i + u_{it}$$

固定効果モデル推定の特徴

- 個体固有効果 Z_i が確定的であると仮定されている
- ダミー変数を利用する (Within推定は次式の b のみを先に推定し、その後、ある制約情報を利用して各 d_i を求める)

$$Y_{it} = a + bX_{it} + d_1D_1 + d_2D_2 + \cdots + u_{it}$$

(5)

固定効果推定

- 人種や学歴など時間が経過しても変化しない変数は次式でキャンセルアウトされる

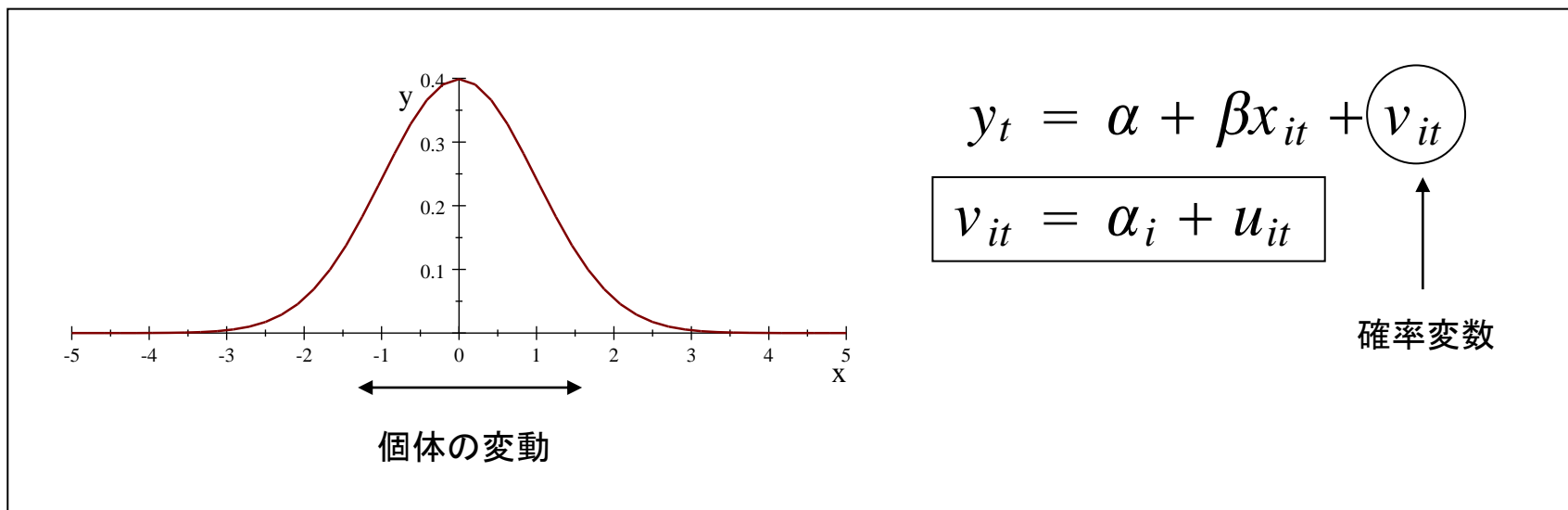
$$(Y_{it} - \bar{Y}_i) = (X_{it} - \bar{X}_i)b + (\epsilon_{it} - \bar{\epsilon}_i)$$

重要: 固定効果モデル推定の特徴

□ Time Invariantな変数は利用できない。

ランダム効果モデル

- 個体ごとの独自の変動を確率変数として考えます。



ランダム効果モデル

■ 賃金に関するモデルの作成(nls_panel.wf1)

id = identifier for panel individual; 716 total
year = year interviewed (1982, 1983, 1985, 1987, 1988)
lwage = $\ln(\text{wage}/\text{GNP deflator})$
educ = current grade completed
south = 1 if south
black = 1 if black; 0 if white
union = 1 if union member
exper = total work experience
exper2 = exper^2
tenure = job tenure, in years
tenure2 = tenure^2

目的: 所得を被説明変数とする回帰モデルを作成すること。

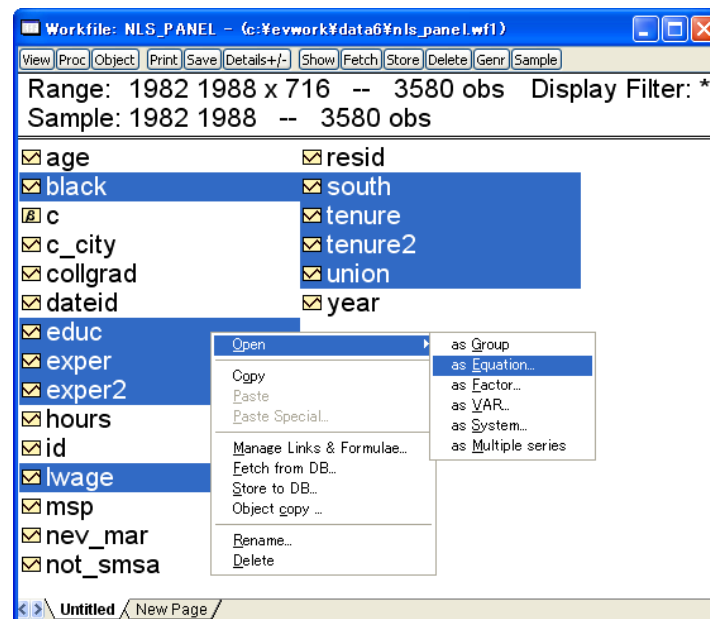
変数の多いモデル

■ ワークファイルウィンドウを利用します

操作1: 次の順番で変数を選択し、右クリックしてOpen/As Equation...と操作します。

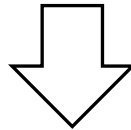
lwage exper exper2 tenure tenure2 south union educ black

操作2: 被説明変数lwageの次に定数項のcを入力してプーリング推計を実行します。オブジェクト名はEQ01とします。



固定効果モデル

操作:EQ01でクロスセクション間の固定効果モデルを推定を実行します。



時間が経過しても変化しない(time invariant)変数が含まれていることが原因です。

ランダム効果モデル

操作1: eq01をオブジェクトコピーして「eq01r」を作成し、ランダム効果モデルを推定します。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.533929	0.079722	6.697408	0.0000
EXPER	0.043617	0.006345	6.874478	0.0000
EXPER2	-0.000561	0.000262	-2.140427	0.0324
TENURE	0.014154	0.003160	4.478901	0.0000
TENURE2	-0.000755	0.000194	-3.886833	0.0001
SOUTH	-0.081812	0.022366	-3.657897	0.0003
UNION	0.080235	0.013187	6.084619	0.0000
EDUC	0.073254	0.005320	13.76943	0.0000
BLACK	-0.116737	0.030148	-3.872140	0.0001
Effects Specification			S.D.	Rho
Cross-section random			0.329050	0.7399
Idiosyncratic random			0.195110	0.2601

◆ランダム効果モデルの場合、時間変化しない変数も利用できます。

◆すべてのパラメータが有意になりました。

誤差

■ ランダム効果モデルの誤差の情報

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.533929	0.079722	6.697408	0.0000
EXPER	0.043617	0.006345	6.874478	0.0000
EXPER2	-0.000561	0.000262	-2.140427	0.0324
TENURE	0.014154	0.003160	4.478901	0.0000
TENURE2	-0.000755	0.000194	-3.886833	0.0001
SOUTH	-0.081812	0.022366	-3.657897	0.0003
UNION	0.080235	0.013187	6.084619	0.0000
EDUC	0.073254	0.005320	13.76943	0.0000
BLACK	-0.116737	0.030148	-3.872140	0.0001
Effects Specification				
			S.D.	Rho
Cross-section random			0.329050	0.7399
Idiosyncratic random			0.195110	0.2601

個体間の誤差の分布 σ_v :0.329050

個体内の誤差の分布 σ_ε :0.195110

分散比:ここでは個体間の変動の方が大きい

$$0.7399 = \frac{0.32905^2}{0.32905^2 + 0.19511^2}$$

$$\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_v^2 + \sigma_\varepsilon^2}}$$

この θ を利用して、ランダム効果を計算します。

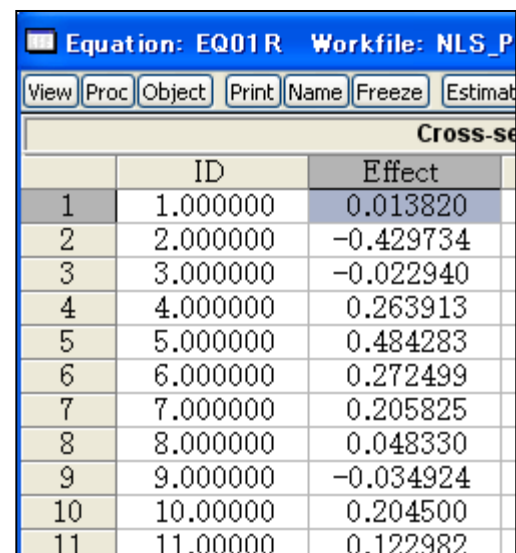
ランダム効果

■ 個々のランダム効果を見てみましょう。

操作: eq01rでView/Fixed,Random Effects/Cross-section Effectsと操作します。

ランダム効果も固定効果と同じく、観測できない個体の異質性です。ランダム効果の場合、個体の変数との共分散(相関)はゼロとして、考えます。

ランダム効果モデルの変量効果は**一**致性を有します



	ID	Effect
1	1.000000	0.013820
2	2.000000	-0.429734
3	3.000000	-0.022940
4	4.000000	0.263913
5	5.000000	0.484283
6	6.000000	0.272499
7	7.000000	0.205825
8	8.000000	0.048330
9	9.000000	-0.034924
10	10.00000	0.204500
11	11.00000	0.122982

ハウスマン検定

操作1: ランダム効果モデルを推定したeq01rでView/Fixed/Random Effect Testing/Correlated Random Effects – Hausman Testと操作します。

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.	
Cross-section random	20.437076	6	0.0023	
Cross-section random effects test comparisons:				
Variable	Fixed	Random	Var(Diff.)	Prob.
EXPER	0.041083	0.043617	0.000004	0.1798
EXPER2	-0.000409	-0.000561	0.000000	0.0504
TENURE	0.013909	0.014154	0.000001	0.7782
TENURE2	-0.000896	-0.000755	0.000000	0.0380
SOUTH	-0.016322	-0.081812	0.000807	0.0211
UNION	0.063697	0.080235	0.000029	0.0022

帰無仮説: $Cov(x_{it}, v_i) = 0$

対立仮説: $Cov(x_{it}, v_i) \neq 0$

P値が0.23%なので、帰無仮説(ランダム効果)を棄却し、対立仮説の固定効果モデルを採用します。

ハウスマン検定の結果、固定効果モデルを支持する結果となりました。ここでは時間変化しない変数(EDUCとBLACK)は自動的にモデルから削除されています

ハウスマン検定の結果画面

操作1: ハウスマン検定の結果の画面を下の方にスクロールしてください。

EDUCとBLACKは時間が経過しても変化しません！

操作2: EQ01をオブジェクトコピーしてEQ01fを作成し、EDUCとBKACKを削除して固定効果モデルを推定します。

Sample: 1982 1988
Periods included: 5
Cross-sections included: 716
Total panel (balanced) observations: 3580
WARNING: estimated coefficient covariance matrix is of reduced rank

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.450034	0.040140	36.12443	0.0000
EXPER	0.041083	0.006620	6.205904	0.0000
EXPER2	-0.000409	0.000273	-1.496532	0.1346
TENURE	0.013909	0.003278	4.243324	0.0000
TENURE2	-0.000896	0.000206	-4.353571	0.0000
SOUTH	-0.016322	0.036149	-0.451531	0.6516
UNION	0.063697	0.014254	4.468790	0.0000
EDUC	NA	NA	NA	NA
BLACK	NA	NA	NA	NA

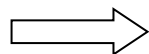
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.450034	0.040140	36.12443	0.0000
EXPER	0.041083	0.006620	6.205904	0.0000
EXPER2	-0.000409	0.000273	-1.496532	0.1346
TENURE	0.013909	0.003278	4.243324	0.0000
TENURE2	-0.000896	0.000206	-4.353571	0.0000
SOUTH	-0.016322	0.036149	-0.451531	0.6516
UNION	0.063697	0.014254	4.468790	0.0000

左図のEQ01fが適切なモデルだと考えられます。

操作の流れ

- 固定効果モデルとランダム効果モデルの適正を判別します。

ランダム効果
モデルの推定



ハウスマン検定

*ハウスマン検定を行う際、準備として固定効果モデルの推定を行う必要はありません。

ハウスマン検定

「ランダム効果モデルで確率変数と説明変数は無相関である」という仮定について調べるための検定です。これらが相関していれば、ランダム効果モデルの採用は不適切である、ということになります。

ランダム効果推定量

v_i が個体に対して確率的である時、一般化最小二乗法を用いて次式を推定します。

$$(Y_{it} - \theta \bar{Y}_i) = (1 - \theta)a + (X_{it} - \theta \bar{X}_i)b + \{(1 - \theta)v_i + (\epsilon_{it} - \theta \bar{\epsilon}_i)\} \quad (4)$$

一般化最小二乗法による推定量で、変数 z を次のようにGLS変換します。 e は個体内、 u は個体間の残差変動を示す。

$$z_{it}^* = z_{it} - \hat{\theta}_i \bar{z}_i$$

$$\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{T_i \hat{\sigma}_u^2 + \hat{\sigma}_e^2}}$$

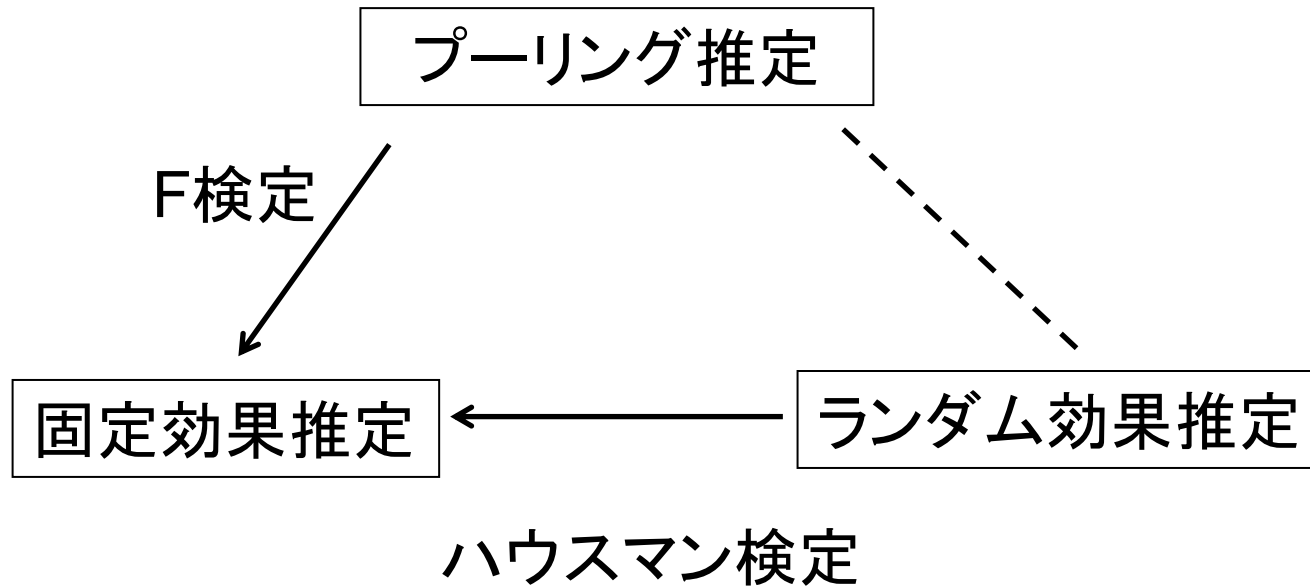
Hausman(1978)検定

- ランダム効果モデルの推定量を有効性があるとし、一方の固定効果モデルは一致性のみあるとします。
- ハウスマン検定統計量はカイ二乗分布します。

$$H = (\beta_c - \beta_e)'(V_c - V_e)^{-1}(\beta_c - \beta_e)$$

H0: ランダム効果モデルの定式化に誤りはない。

モデルの選択



Wooldridge(2002)

- 職業訓練の補助金(1987-1989)が、企業の製品廃棄率に与える影響を分析

被説明変数 lscrap (製品廃棄率)

説明変数 d88 年次ダミー
 d89 年次ダミー
 union 労働組合あり
 grant 補助金受給
 grant_1 grantのラグ項

操作: ワークファイル jtrain.wf1を開きます。

Wooldridge(2002)

操作1:クロスセクションIDをfcode, 時間変数をyearとしてパネルデータの設定を行います。

操作2:次のモデルEQ01を固定効果モデルとして推定します。

$$lscrap_{it} = \alpha_0 + \alpha_i + \beta_1 d88_{it} + \beta_2 d89_{it} + \beta_3 grant_{it} + \beta_4 grant_{it-1} + e_{it}$$

但し、誤差項の系列相関を考慮して、Panel Optionsタブにある係数共分散の項目で、White periodを選び、さらに、自由度調整のオプションをチェックします。

The screenshot shows the 'Panel Options' tab of a software interface. It contains three main sections: 'Effects specification', 'Weights', and 'Coef covariance method'. In the 'Effects specification' section, 'Cross-section' is set to 'Fixed' and 'Period' is set to 'None'. In the 'Weights' section, 'GLS Weights' is set to 'No weights'. In the 'Coef covariance method' section, 'White period' is selected from a dropdown menu, and the checkbox for 'No d.f. correction' is checked. Red arrows point to the 'Cross-section' dropdown, the 'White period' dropdown, and the 'No d.f. correction' checkbox.

Wooldridge(2002)

Sample: 1987 1989
Periods included: 3
Cross-sections included: 54
Total panel (balanced) observations: 162
White period standard errors & covariance (no d.f. correction)
WARNING: estimated coefficient covariance matrix is of reduced rank

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.597434	0.062489	9.560565	0.0000
D88	-0.080216	0.095719	-0.838033	0.4039
D89	-0.247203	0.192514	-1.284075	0.2020
GRANT	-0.252315	0.140329	-1.798022	0.0751
GRANT_1	-0.421589	0.276335	-1.525648	0.1301

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.927572	Mean dependent var	0.393681
Adjusted R-squared	0.887876	S.D. dependent var	1.486471
S.E. of regression	0.497744	Akaike info crite...	1.715383
Sum squared resid	25.76593	Schwarz criterion	2.820819
Log likelihood	-80.94602	Hannan-Quinn criter.	2.164207
F-statistic	23.36680	Durbin-Watson stat	1.996983
Prob(F-statistic)	0.000000		

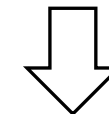
固定効果モデル

+

Robust Option

+

時点数<個体数



Reduced Rankの警告

係数共分散行列

標準誤差の計算に関係します

White cross-section...誤差項について
各時点での同時相関を考慮する。同時
相関と不均一分散に対して堅牢である。

Ordinary
White cross-section
White period
White (diagonal)
Cross-section SUR (PCSE)
Cross-section weights (PCSE)
Period SUR (PCSE)
Period weights (PCSE)

$$\left(\frac{N^*}{N^*-K^*}\right)(\sum_t X_t' X_t)^{-1}(\sum_t X_t' \hat{\epsilon}_t \hat{\epsilon}_t' X_t)(\sum_t X_t' X_t)^{-1}$$

$$\begin{array}{cc} ID_1 & ID_2 \\ t_1 & e_{11} & e_{21} \\ \hline t_2 & e_{12} & e_{22} \end{array} \rightarrow$$

係数共分散行列

標準誤差の計算に関係します

White period... 個体ごとの不均一分散と時系列方向の系列相関を考慮。

Ordinary
White cross-section
White period
White (diagonal)
Cross-section SUR (PCSE)
Cross-section weights (PCSE)
Period SUR (PCSE)
Period weights (PCSE)

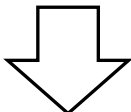
$$\left(\frac{N^*}{N^* - K^*} \right) (\sum_i X_i' X_i)^{-1} (\sum_i X_i' \hat{\epsilon}_i \hat{\epsilon}_i' X_i) (\sum_i X_i' X_i)^{-1}$$

	ID_1	ID_2
t_1	e_{11}	e_{21}
t_2	e_{12}	e_{22}

↓

*各選択肢の詳細は
PDFマニュアルの
Pooled Estimationの項
を参照。

Wooldridge(2002)

$$lscrap_{it} = \alpha_0 + \alpha_i + \beta_1 d88_{it} + \beta_2 d89_{it} + \beta_3 grant_{it} + \beta_4 grant_{it-1} + u_{it} \quad (A)$$


$$d(lscrap_{it}) = \alpha + \beta_1 d89_{it} + \beta_2 d(grant_{it}) + \beta_3 d(grant_{it-1}) + e_{it} \quad (B)$$

B式を単純な(効果の無い)回帰モデルとして推定し、 e_{it} に1次の自己相関が無い(系列相関が無い)ことを確認する。

Wooldridge(2002)

$$d(lscrap_{it}) = \alpha + \beta_1 d89_{it} + \beta_2 d(grant_{it}) + \beta_3 d(grant_{it-1}) + e_{it}$$

操作: 定式を効果を考慮しないモデルとしてEQ02を推定する。推定オプションとしてWhite periodを利用する。

Coef covariance method

White period

☒ No d.f. correction

Sample (adjusted): 1988 1989

Periods included: 2

Cross-sections included: 54

Total panel (balanced) observations: 108

White period standard errors & covariance (no d.f. correction)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.090607	0.088082	-1.028671	0.3060
D89	-0.096208	0.111002	-0.866721	0.3881
D(GRANT)	-0.222781	0.128580	-1.732624	0.0861
D(GRANT_1)	-0.351246	0.264662	-1.327147	0.1874

Wooldridge(2002)

操作1: 残差に系列相関があるか確認します。パネルデータのモデルではGMM推定以外では、検定機能はありませんので、手作業で操作します。

EQ02でProc/Make Residual Series...として残差resid01を取り出します。

操作2: 残差の自己回帰モデルEQ03を推定します。

$$resid01_{it} = \gamma resid01_{it-1} + v_{it}$$

*残差に1次の自己相関が無い時、 $\gamma = -0.5$ となります。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	0.236906	0.133357	1.776481	0.0814

Wooldridge(2002)

操作1:EQ03でView/Coefficient Diagnostics/Wald Coefficient Restrictions...と操作し、 $c(1)=-0.5$ として、検定を実行します。

Wald Test: Equation: Untitled			
Test Statistic	Value	df	Probability
t-statistic	5.525812	53	0.0000
F-statistic	30.53460	(1, 53)	0.0000
Chi-square	30.53460	1	0.0000
Null Hypothesis: C(1)=-0.5 Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
0.5 + C(1)	0.736906	0.133357	
Restrictions are linear in coefficients.			

帰無仮説は棄却されるので、B式に系列相関が存在します。

Breush and Pagan(1980)

■ プーリング推定とランダム効果推定の選択

$$y_{it} = \alpha + x_{it}\beta + v_{it}$$

この式をOLS推定して、

$$\lambda_{LM} = \frac{(n\bar{T})^2}{2} \left(\frac{A_1^2}{\left(\sum_i T_i^2\right) - n\bar{T}} \right)$$

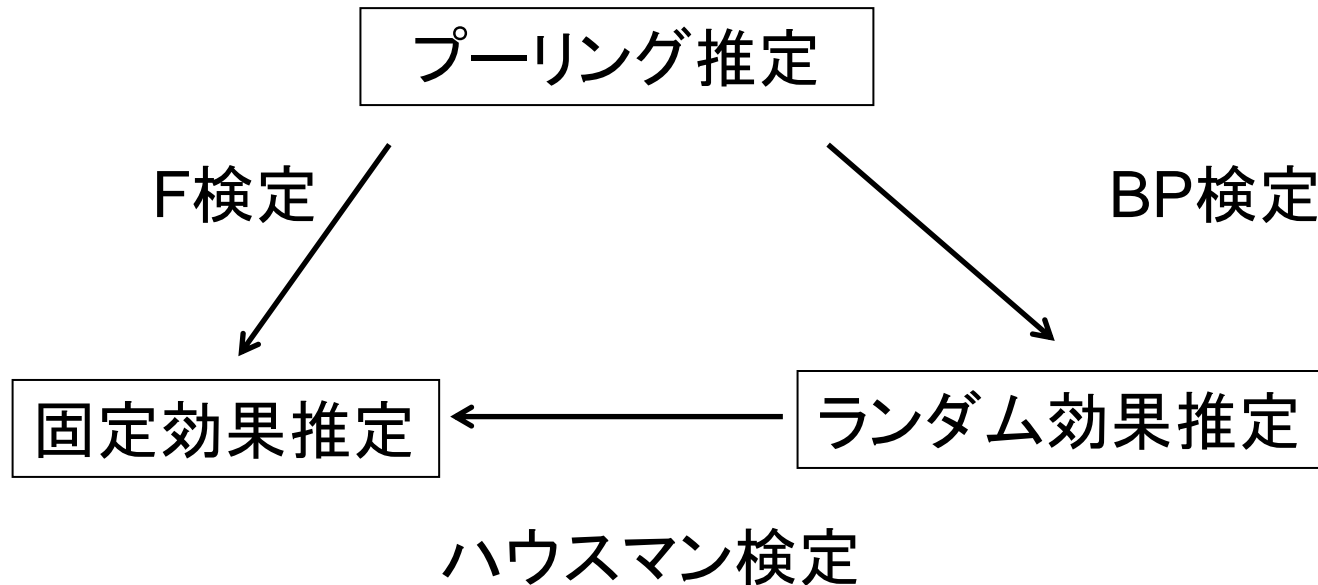
$$A_1 = 1 - \frac{\sum_{i=1}^n \left(\sum_{t=1}^{T_i} v_{it} \right)^2}{\sum_i \sum_t v_{it}^2}$$

Breusch and Pagan(1980)

- 帰無仮説は「モデルはプーリング推定が適している」

$$y_{it} = \alpha + x_{it}\beta + v_{it}$$

モデルの選択



Grunfeld(1958)

シンプルなプーリングモデルにランダム効果を設定するべきか?
1935-1954年(20年)の企業の投資関数を推定します。サンプルデータはgrunfeld_baltagi_panel.wf1

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

I:投資金額

F:企業価値(発行済み株式の価値)

C:資本ストック(変数はc01)

操作1:プーリングモデルeq01を推定します。

i c f c01

Grunfeld(1958)

Method: Panel Least Squares
Date: 06/29/15 Time: 08:59
Sample: 1935 1954
Periods included: 20
Cross-sections included: 10
Total panel (balanced) observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-42.71437	9.511676	-4.490730	0.0000
F	0.115562	0.005836	19.80259	0.0000
C01	0.230678	0.025476	9.054808	0.0000

操作:このモデルにランダム効果を追加すべきか、検定します。
View/Fixed-Random Effects Testing/Omitted Random Effects –
Lagrange Multiplierと操作します。

Grunfeld(1958)

帰無仮説
「プーリング
モデルが良い」

Lagrange Multiplier Tests for Random Effects

Null hypotheses: No effects

Alternative hypotheses: Two-sided (Breusch-Pagan) and one
-sided (all others) alternatives

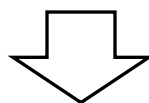
	Cross-section	Test Hypothesis Time	Both
Breusch-Pagan	798.1615 (0.0000)	6.453882 (0.0111)	804.6154 (0.0000)
Honda	28.25175 (0.0000)	-2.540449 --	18.18064 (0.0000)
King-Wu	28.25175 (0.0000)	-2.540449 --	21.83221 (0.0000)
Standardized Honda	32.66605 (0.0000)	-2.432565 --	16.29814 (0.0000)
Standardized King-Wu	32.66605 (0.0000)	-2.432565 --	20.96591 (0.0000)
Gourieriou, et al.*	--	--	798.1615 (< 0.01)

*Mixed chi-square asymptotic critical values:

1%	7.289
5%	4.321
10%	2.952

クロスセクション方向の相関

パネルデータではクロスセクション方向の誤差項は独立であるという仮定があります。



仮定が満たされないと、推定量の効率性は失われ、推定後の検定統計量にバイアスが生じます。

Pesaran(2004)

$$y_{it} = \beta_i' x_{it} + u_{it}$$

帰無仮説:クロスセクション方向に依存性なし

$$H_0 : \rho_{ij} = \text{Corr}(u_{it}, u_{jt}) = 0 \text{ for } i \neq j$$

Pesaran(2004)

相関係数の定義

バランスパネルの場合：

$$\hat{\rho}_{ij} = \frac{\sum \hat{u}_{it}\hat{u}_{jt}}{(\sum \hat{u}_{it}^2)^{1/2}(\sum \hat{u}_{jt}^2)^{1/2}}$$

アンバランスパネルの場合：

$$\hat{\rho}_{ij} = \frac{\sum (\hat{u}_{it}-\tilde{u}_i)(\hat{u}_{jt}-\tilde{u}_j)}{\sqrt{\sum (\hat{u}_{it}-\tilde{u}_i)^2} \sqrt{\sum (\hat{u}_{jt}-\tilde{u}_j)^2}}$$

ここで、

$$\tilde{u}_i = \frac{\sum \hat{u}_{it}}{T_{ij}}$$

クロスセクション方向の相関

操作:EViewsワークファイルgasoline.wf1を開きます。

Baltagi(2008):

OECD加盟18カ国のガソリン消費量データ

LGASPCAR:1台あたりのガソリン消費量の対数値

LINCOME_P:一人当たり実質所得の対数値

LRPMG:実質ガソリン価格の対数値

LCARPCAP:一人当たり自動車所有台数

$$LGASPCAR_{it} = \alpha_0 + \alpha_i + \beta_1 LINCOME_{it} + \beta_2 LRPMG_{it} + \beta_3 LCARPCAP_{it} + u_{it}$$

Baltagi(2008)

$$LGASPCAR_{it} = \alpha_0 + \alpha_i + \beta_1 LINCOME_{it} + \beta_2 LRPMG_{it} + \beta_3 LCARPCAP_{it} + u_{it}$$

操作:上記のガソリンの消費モデルEQ01を、固定効果モデルとして推定します。

Sample: 1960 1978 Periods included: 19 Cross-sections included: 18 Total panel (balanced) observations: 342				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.402670	0.225309	10.66387	0.0000
LINCOME	0.662250	0.073386	9.024191	0.0000
LRPMG	-0.321702	0.044099	-7.294964	0.0000
LCARPCAP	-0.640483	0.029679	-21.58045	0.0000
Effects Specification				
Cross-section fixed (dummy variables)				

Baltagi(2008)

操作:EQ01において、View/Residual Diagnostics/Cross-section Dependence Testと操作して、クロスセクション方向の相関についての検定を実行します。

Residual Cross-Section Dependence Test			
Null hypothesis: No cross-section dependence (correlation) in residuals			
Equation: EQ01			
Periods included: 19			
Cross-sections included: 18			
Total panel observations: 342			
Cross-section effects were removed during estimation			
Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	1027.135	153	0.0000
Pesaran scaled LM	48.94200		0.0000
Bias-corrected scaled LM	48.44200		0.0000
Pesaran CD	3.249563		0.0012

帰無仮説: 残差にクロスセクション方向の相関はない。

Pesaran(2008)

- ❑ Breusch-Pagan LM :4つの検定統計量のうち、もっとも一般的な統計量。
- ❑ Pesaran Scaled LM:データ数 N が大きいときは、このPesaran Scaled LMの方が良い。ただし、 N が小さい場合は検定統計量が正規分布しないので注意。
- ❑ Pesaran CD: N と T が小さい場合でも優れた統計量を提供する。
- ❑ バイアス修正のLM統計量: Scaled LMのバイアス修正版。固定効果モデルの場合にのみ出力。

パネルデータの分析

- 世界208の国と地域に関する一人当たりのGDPと、人口増加率の関係を分析します(pwt61extract.wf1)。

被説明変数 y : 一人当たりの実質GDPの対米国比
説明変数 $dlog(pop)$: 人口増加率

国や地域の特徴を固定効果として考えるべきか?

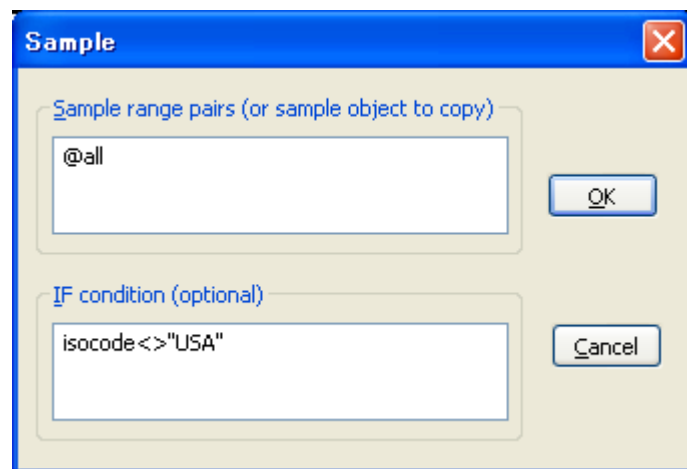
除外

- 米国自身のデータが分析対象のyとpopに含まれていますので、最初にサンプルから外します。

操作: Quick/Sampleと操作してif conditionの項目につぎのように入力します。

isocode <> "USA"

条件に利用する個体文字列は大文字と小文字を区別します。小文字の"usa"ではエラーになります。



Range: 1950 2000 x 208 -- 10208 obs
Sample: 1950 2000 if isocode<>"USA" -- 10157 obs

Sampleの表示が変わります。

単位根検定

時系列データのモデル推定については単位根の存在を確認する必要があります。

単純なAR(1)過程のデータ

$$y_t = \rho y_{t-1} + x_t' \delta + \epsilon_t$$

ρ の絶対値が1以上なら非定常で、1未満であれば定常です。通常の時系列データの場合はADF検定統計量など、一つの検定統計量を使って判定を行いましたが、パネル単位根検定の場合、6個の検定統計量を複合的に利用して判定を行います。

基本的な考え方

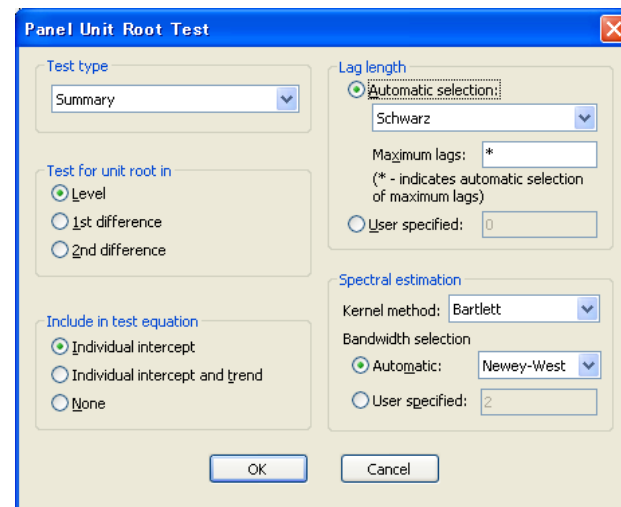
Common...AR過程で同じ係数を利用します。

Individual...AR過程で個体ごとに異なる係数を利用します。

パネル単位根検定

操作1: 被説明変数 y の単位根検定を実行します。

通常の時系列データの場合は簡単にグラフ化してトレンドや切片を確認できましたが、個体数の多いパネルデータの場合に視覚的に確認することはできません。それぞれに、切片が異なるという条件で単位根検定を実行します。



Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t^*	-3.42901	0.0003	152	5552
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.96445	0.0247	152	5552
ADF - Fisher Chi-square	397.555	0.0002	152	5552
PP - Fisher Chi-square	462.508	0.0000	152	5620

どの検定統計量を見ても帰無仮説(単位根を持つ)は棄却できません。

操作2: 同じ要領で $d\log(\text{pop})$ について単位根検定を実行します。

固定効果モデルの推定

操作1: 次の式を効果無しで推定します。式はeq01とします。

y c dlog(pop)

人口増加率の係数は有意であることが示されています。

	Coefficient	Std. Error	t-Statistic	Prob.
C	44.93176	0.546861	82.16302	0.0000
DLOG(POP)	-901.0468	23.83234	-37.80773	0.0000

プーリング推計

操作2: 国と地域による特徴が一人当たりのGDPに影響するものとして固定効果モデルを推定します。eq01をオブジェクトコピーしてeq02を作成し、固定効果モデルを推定します。

人口増加率の係数は1/100になり、有意ではなくなりました。

	Coefficient	Std. Error	t-Statistic	Prob.
C	27.92379	0.219137	127.4263	0.0000
DLOG(POP)	-8.371774	10.57512	-0.791648	0.4286

固定効果

操作3: どちらを採用すべきか、Redundant Testを実行して確認してください。

データの考察

- 回帰式の傾きが有意ではない、という結果を受けて個々のデータを調べることにします。

操作1: yとdlog(pop)のデータを日本(JPN)、イタリア(ITA)、中央アフリカ共和国(CAF)について比較します。サンプルを米国を除くすべての地域から、これら3カ国に変更します。

Range: 1950 2000 x 208 -- 10208 obs

Display Filter:

Sample: 1950 2000 if isocode="CAF" or isocode="JPN" or isocode="ITA" -- 153 obs

人口増加率を比較します

操作2: dlog(pop)をView/Descriptive Statistics & Test/ Stats by Classificationコマンドと操作し、isocodeで階層化します。統計量としてはMeanとObsをチェックします。

*50年間の人口増加率の平均値はかなり大きな差があります。

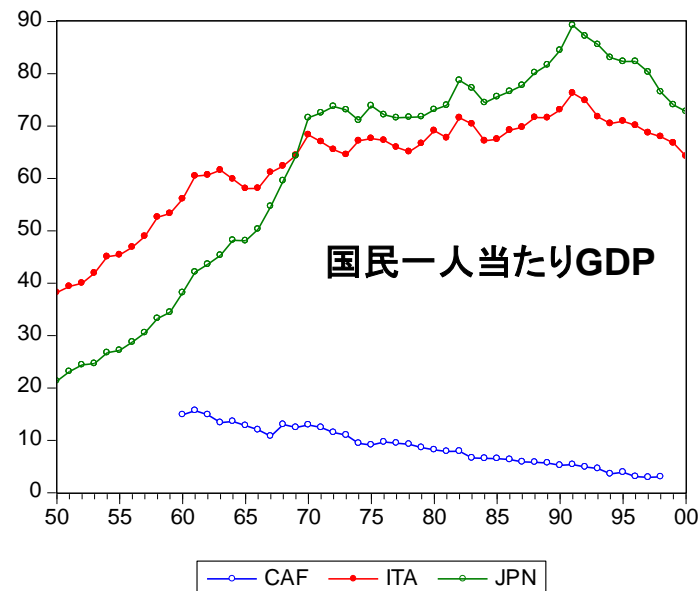
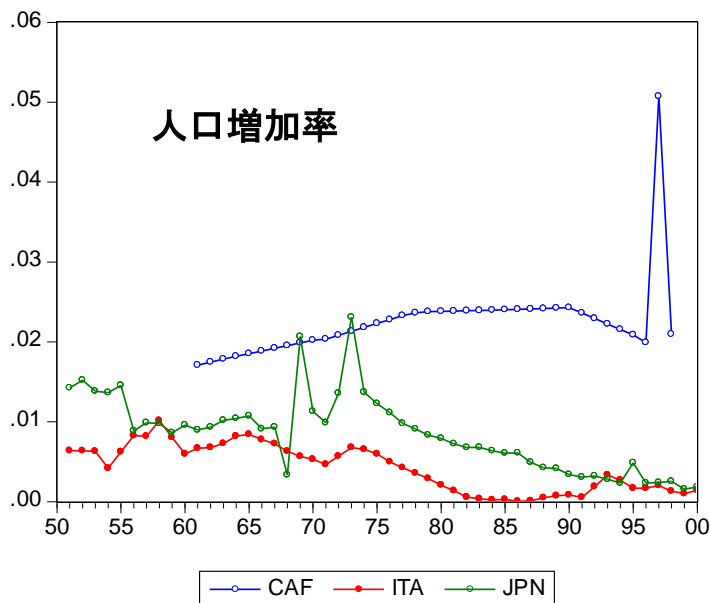
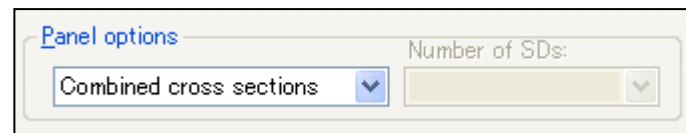
Descriptive Statistics for DLOG(POP)
Categorized by values of ISOCODE
Date: 10/13/07 Time: 07:24
Sample: 1950 2000 IF ISOCODE="...
OR ISOCODE="JPN" OR
ISOCODE="ITA"
Included observations: 138

ISOCODE	Mean	Obs.
CAF	0.022474	38
ITA	0.004197	50
JPN	0.008469	50
All	0.010778	138

グラフの作成

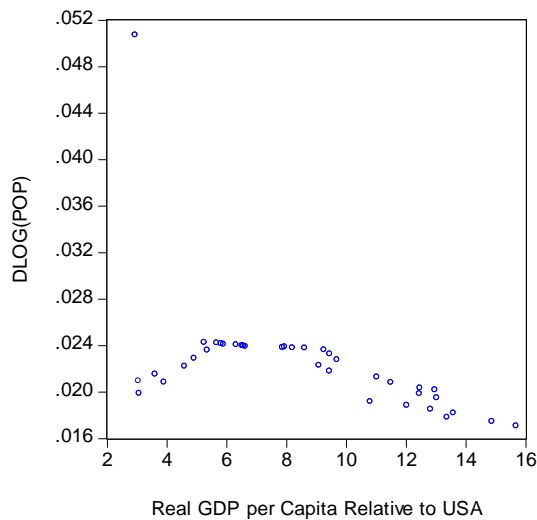
■ 人口増加率と国民一人当たりGDPのグラフ化

操作: $\text{dlog}(\text{pop})$ と y (国民一人当たりGDP)をグラフ化します。Panel optionsの項目はCombined cross sectionsを選択します。

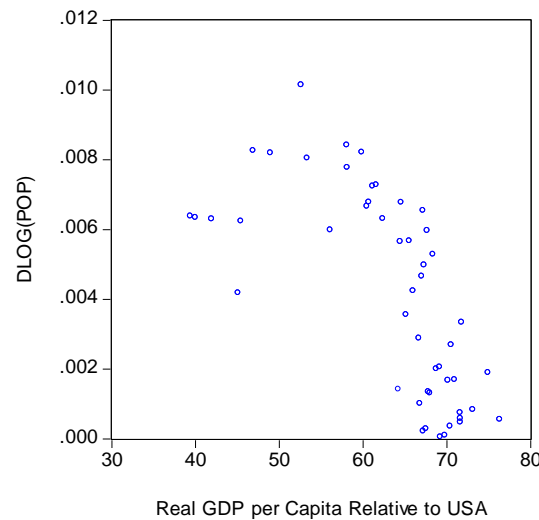


散布図

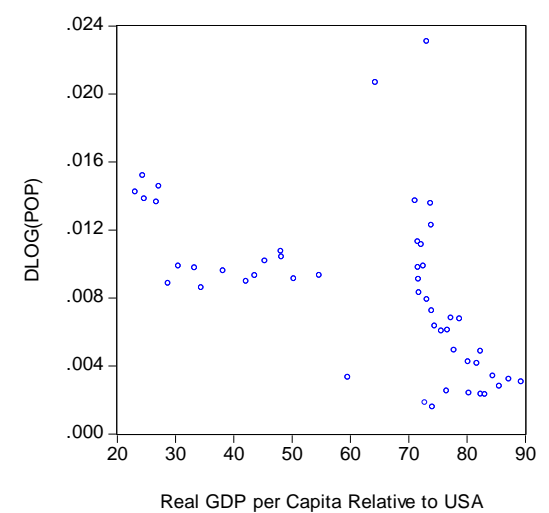
- 人口増加率と国民一人当たりGDPの各国散布図を作成します。OptionにIndividual Cross-Sectionを選択します。



中央アフリカ共和国



イタリア



日本

各国の特徴を固定効果として定式化すると、2変数間の相関を有意な係数で捕らえることができないことがわかりました。

サンプル

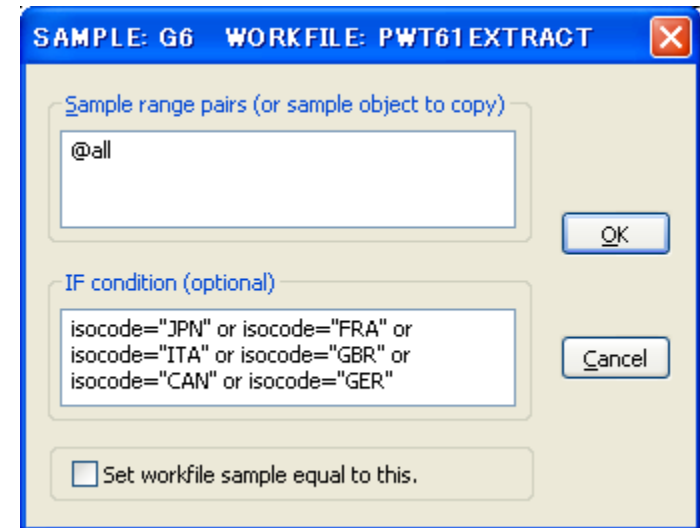
- 先進6カ国にサンプルを変更して、パネル分析を行います。

日本(JPN)、フランス(FRA)、イタリア(ITA)、イギリス(GBR)、カナダ(CAN)、ドイツ(GER)にサンプルを変更します。サンプルの切り替えを簡単に行うために、ここでは最初にサンプルオブジェクトを作成します。

操作1: 新しいオブジェクトとしてサンプルオブジェクトを作成します。オブジェクト名は「g6」とします。

操作2: オブジェクトを作成したら、コマンドウィンドウに次のように入力してサンプルを実際に切り替えます。

smpl g6 ⇄ g6



Range: 1950 2000 x 208 -- 10208 obs Display Filter: *
Sample: 1950 2000 if isocode="JPN" or isocode="FRA" or isocode="IT

固定効果モデル

- 先進6カ国を対象に固定効果モデルで推定します。

操作1: オブジェクト名を「eq_g6」として、固定効果モデルで推定します。
操作2: 固定効果の冗長性をRedundant Testで検定します。

	Coefficient	Std. Error	t-Statistic	Prob.
C	75.59272	1.174237	64.37601	0.0000
DLOG(POP)	-754.0516	141.7995	-5.317729	0.0000

固定効果モデルで推定しても有意な係数を得ました

Effects Test	Statistic	d.f.	Prob.
Cross-section F	35.684935	(5,273)	0.0000
Cross-section Chi-square	140.822284	5	0.0000

固定効果モデルの採用を示す結果になりました。

操作3: ファイル名を「mytest」として保存します。

まとめ

- 一人あたりGDP(米国比)と人口増加率
 - パネル単位根検定
 - 効果なしで推計した場合の係数:-901
 - 固定効果(地域)で推計した場合の係数:-8.37
 - 係数に2桁の違いがあり、後者は有意ではない
 - 先進国と途上国では一人当たりGDPと人口増加率に大きな違いがある
 - 先進6カ国の固定効果モデルの場合:-754

パネルデータにおける推定

パネルデータにおいては不均一分散や系列相関が生じやすいという問題があります。GMMを利用することで、これらの問題を回避できます。

パネルデータにおける自己回帰モデルでは時系列モデルとは異なる特別な問題が発生します。

Arellano and Bond(1991)

ダイナミックパネルデータモデル

■ ラグ項を含むモデル

$$y_{it} = \delta y_{i,t-1} + u_{it} \quad (1)$$

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

固定効果を除去するためにwithin推定を実行します

$$(y_{it} - \bar{y}_i) = \delta(y_{i,t-1} - \bar{y}_{i-1}) + (u_{it} - \bar{u}_i) \quad (2)$$

$(y_{i,t-1} - \bar{y}_{i-1})$ と $(u_{it} - \bar{u}_i)$ に相関が生じてしまう!

ARELLANO and Bond 推定量

1式のwithin推定量ではなく、階差モデルを考える

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (v_{it} - v_{i,t-1}) \quad (3)$$

t=3で具体的に書くと、

$$y_{i3} - y_{i2} = \delta(y_{i2} - y_{i1}) + (v_{i3} - v_{i2}) \quad (4)$$

y_{i1} が操作変数として利用できる

t=4の場合は、

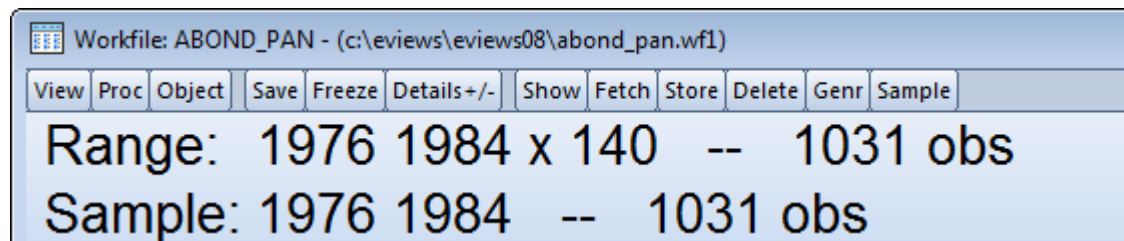
$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (v_{i4} - v_{i3}) \quad (5)$$

y_{i1}, y_{i2} が操作変数として利用できる

GMM

操作1:EViewsのサンプルデータ、Abond_pan.wf1を開きます。このファイルは第41章のサンプルデータとしてEViewsフォルダに入っています。

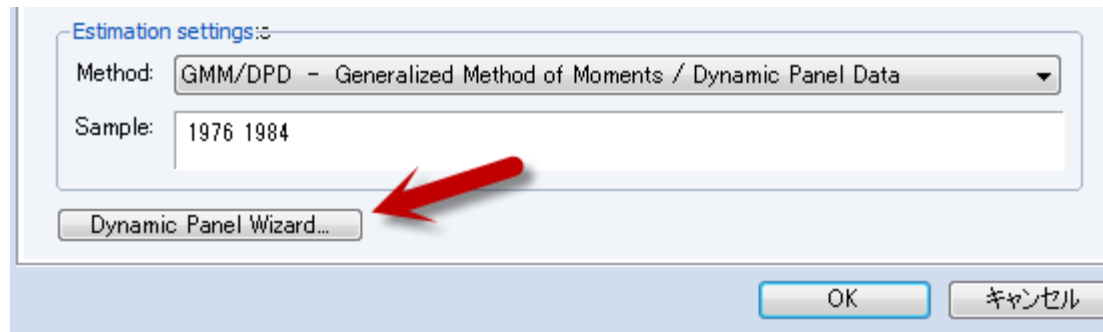
1976年-1984年までの140の企業のデータ
N:企業の従業員数の対数値(被説明変数)
W:実質賃金の対数値
K:資本ストックの対数値
YS:工業生産の対数値



操作2:Quick/Estimate Equationと操作し、GMM/DPDを選択します。パネルデータの場合のみDPDという表示が表れます。

ダイナミックパネル

操作2: Quick/Estimate Equationと操作し、GMM/DPDを選択します。パネルデータの場合のみDPDという表示が表れます。



操作3: ダイアログの下側にあるDynamic Panel Wizard...というボタンをクリックします。

ダイナミックパネル

操作4: Welcomeダイアログで「次へ」をクリックし、被説明変数の設定画面を表示します。

Dynamic Panel Data Model Wizard

Specify dependent variable - Step 1 of 6

Dynamic panel data models have the feature that lags of the dependent variable appear as regressors.

What is your dependent variable?

The specification in this wizard uses lags (from one up to a specified maximum) of the dependent variable as regressors. How many lags of the dependent variable do you want to use as regressors?

From one up to lag(s)

< 戻る(B) 次へ(N) > キャンセル

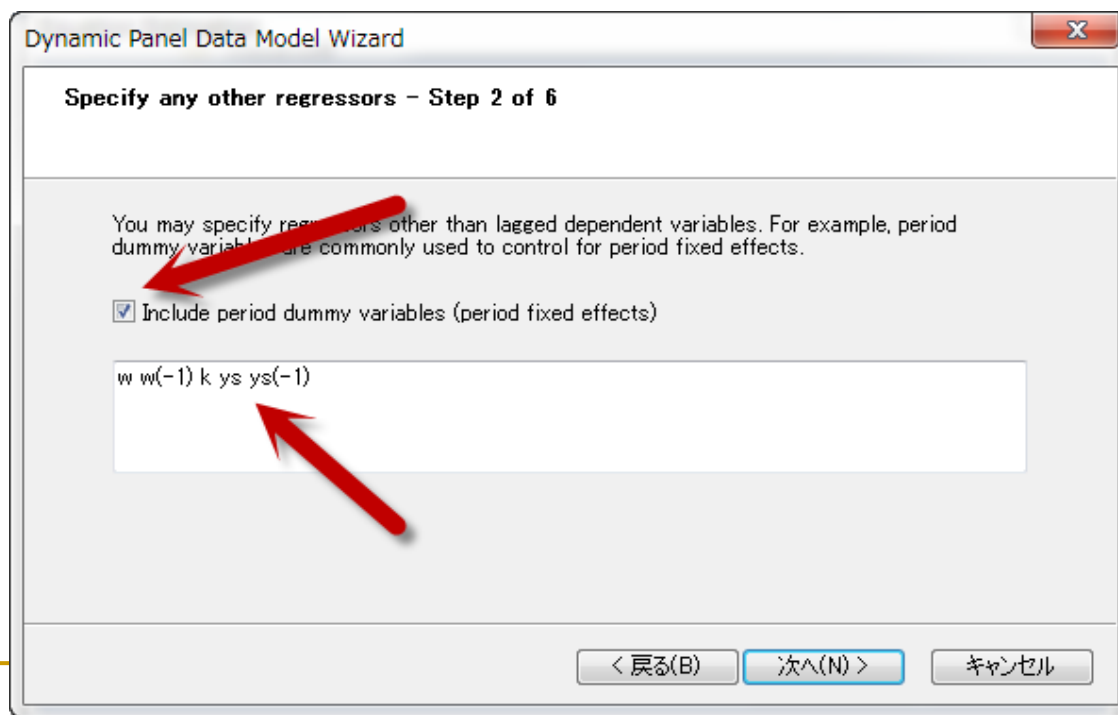
操作5: 被説明変数に「n」、ダイナミックモデルの説明変数として利用するラグ項の次数を「2」として、「次へ」をクリックします。

ダイナミックパネル

操作6: 説明変数を次のように入力します。

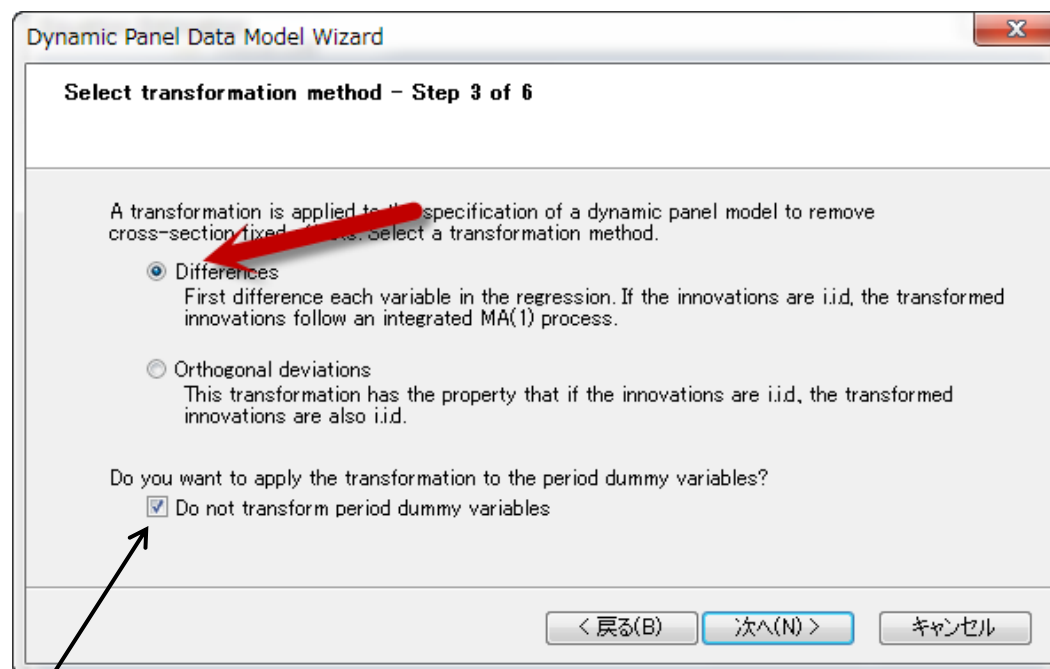
w w(-1) k ys ys(-1)

操作7: 時点ダミーを利用します。Include period dummy variables をチェックして「次へ」をクリックします。



ダイナミックパネル

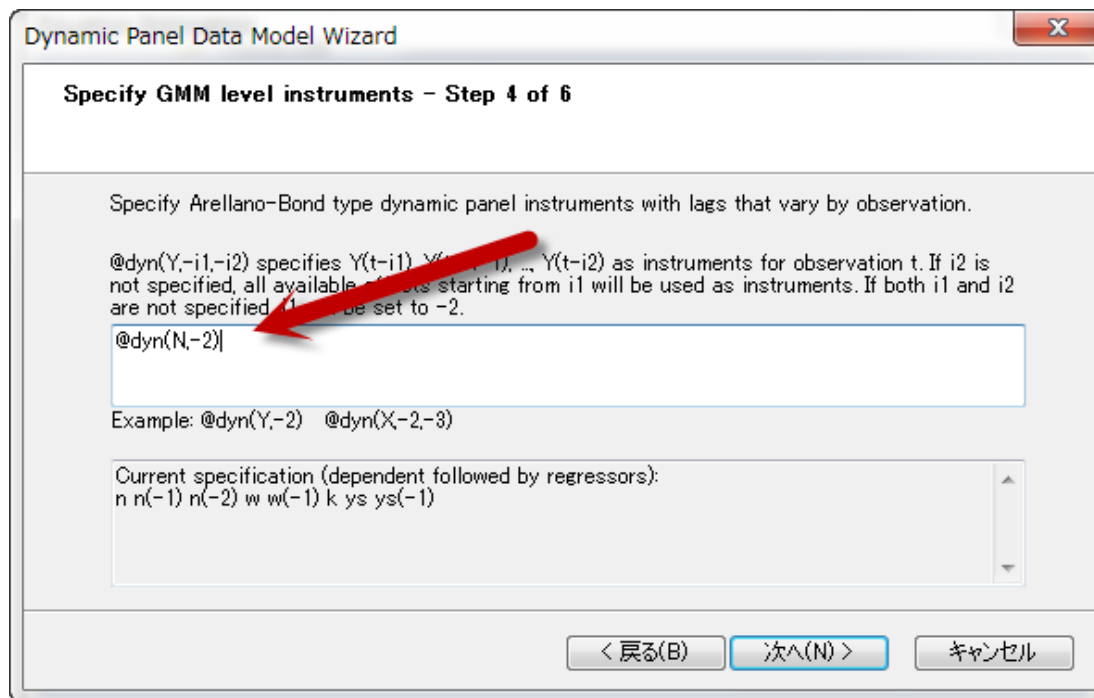
操作8:「Difference」を選択し、クロスセクション方向の固定効果を削除します。そして「次へ」をクリックします。



時点ダミーを選択した場合、自動的に「ダミー変数は変換しない」というオプションのチェックを付けます。

ダイナミックパネル

操作9: 時点について操作変数を設定します。ここではnのラグ項をそのまま、操作変数とします。



時点ダミーを選択した場合、自動的に「ダミー変数は変換しない」というオプションのチェックを付けます。

ダイナミックパネル

操作10: 一般的な操作変数を設定します。ここでは階差を取る選択をしていますので、左側のダイアログに入力し、「次へ」をクリックします。

c w w(-1) k ys ys(-1)

Dynamic Panel Data Model Wizard

Specify regular instruments - Step 5 of 6

List all other instruments in one of the following edit boxes, depending on whether you wish to apply the transformation to the instrument.

Transform (differences):
c w w(-1) k ys ys(-1)|

No transformation:

Note: the period dummies will automatically be added to the no-transformation list.

Current specification (dependent followed by regressors):
n n(-1) n(-2) w w(-1) k ys ys(-1)

< 戻る(B) 次へ(N) > キャンセル

通常はダイアログ下にある情報を入力します。

ダイナミックパネル

操作11:加重行列の設定を行います。ここは難しい所です。ここでは変更せずに「次へ」をクリックします。

Dynamic Panel Data Model Wizard

Select estimation method - Step 6 of 6

Dynamic panel data models are estimated by GMM. Specify the number of iterations, the weighting matrix, and a method to compute standard errors.

GMM iterations: 2-step (update weights once)

GMM weighting matrix:

☐ Period SUR
(innovations have same time series correlation structure for all cross-sections)

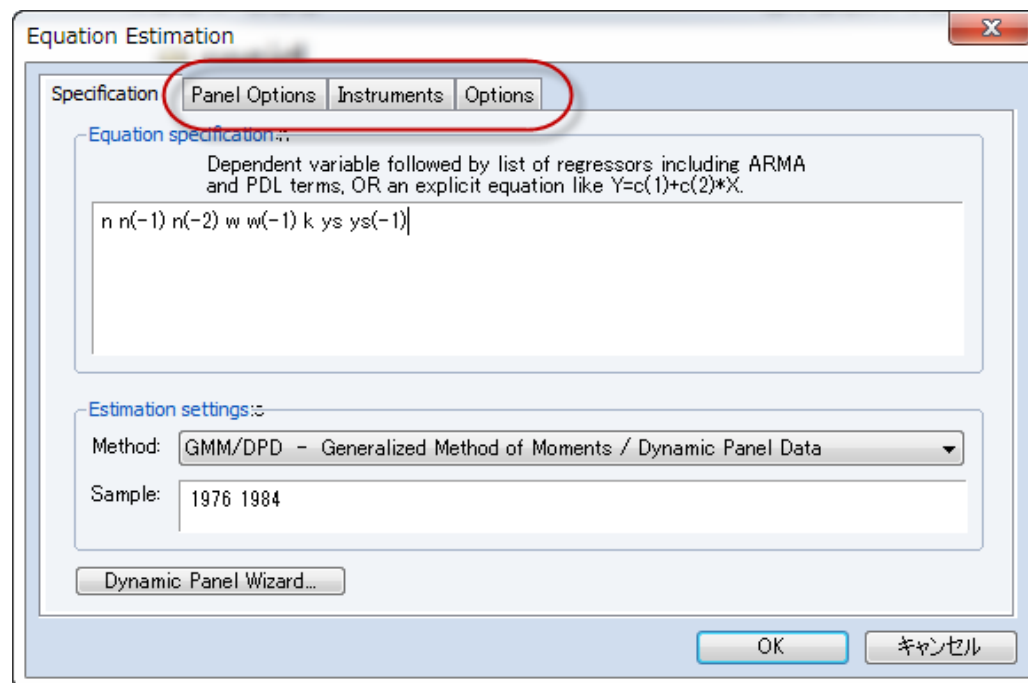
☒ White period
(innovations have time series correlation structure that varies by cross-section)

Standard errors:
Robust - White period weights from final iteration

< 戻る(B) 次へ(N) > キャンセル

ダイナミックパネル

操作12: すべての操作が完了すると、選択項目を反映した推定のダイアログを表示します。



操作13: 各オプションタブの内容を確認したら、OKボタンをクリックし、推定を実行します。

サーガンテスト

パネルデータによるモデル推定のEquationオブジェクトに用意されている検定機能は系列相関の検定だけで、外生性に関するものは無い。

操作1:手作業でサーガンテスト(操作変数の外生性の検定)を実行します。

```
scalar pval=@chisq(30.11247,25)
```

自由度=操作変数のランク(個数)-パラメータの個数

帰無仮説:採用した過剰識別の状態の操作変数は適切である。

参考文献

- Stock and Watson (2015) “Introduction to Econometrics”
- 縄田和満、「EViewsによる計量経済分析入門」、朝倉書店
- 松浦克己、コリン・マッケンジー「EViewsによる計量経済分析」、東洋経済新報社
- 山本 拓、「計量経済学」、新世社