シリーズ

応用計量経済学⑽

横浜市立大学商学部　松浦　克己
大阪大学国際公共政策研究科　Colin McKenzie

第10章　パネル分析

1　パネルデータとその意義

（パネルデータ）

都道府県別のデータは一時点では47である。このようなデータをクロスセクションデータといった。また20年間にわたる東京都のデータを時系列データといった。47都道府県のデータを20年間集めたようなデータをクロスセクションと時系列データのプーリング・データ（pooling cross-section and time-series data）あるいはパネルデータ（panel data）という。

より一般的にいえば同一経済主体について複数時点のデータを集めたものをパネルデータという。売上げと設備投資の関係を検証しようとして、上場企業の財務データが、

パネルA	パネルB
新日鐵(91年)売上げ設備投資	新日鐵(91年)売上げ設備投資
NKK(91年)売上げ設備投資	新日鐵(92年)売上げ設備投資
川崎製鉄(91年)売上げ設備投資	………
住友金属(91年)売上げ設備投資	新日鐵(98年)売上げ設備投資
神戸製鋼(91年)売上げ設備投資	NKK(91年)売上げ設備投資
新日鐵(92年)売上げ設備投資	………
NKK(92年)売上げ設備投資	NKK(98年)売上げ設備投資
川崎製鉄(92年)売上げ設備投資	川崎製鉄(91年)売上げ設備投資
住友金属(92年)売上げ設備投資	………
神戸製鋼(92年)売上げ設備投資	川崎製鉄(98年)売上げ設備投資
………	住友金属(91年)売上げ設備投資
新日鐵(98年)売上げ設備投資	………
NKK(98年)売上げ設備投資	住友金属(98年)売上げ設備投資
川崎製鉄(98年)売上げ設備投資	神戸製鋼(91年)売上げ設備投資
住友金属(98年)売上げ設備投資	………
神戸製鋼(98年)売上げ設備投資	神戸製鋼(98年)売上げ設備投資

というように、５社のデータが８年分あるようなケースである。パネルＡでは時点毎に各経済主体のデータが並べられている。パネルＢでは各経済主体毎にデータが並べられている。

このようにある経済主体ｉについてｔ時点のデータｙ（たとえば売上げ）が得られるとき、それをy_{it}と表記する。このパネルデータを用いた分析をパネル分析（panel analysis）という。

上にあげた例のように全ての経済主体ｉについて全ての時点ｔのデータが揃っているとき、これをバランスしたパネルデータ（balanced panel data）という。ある経済主体iについて一部の時点のデータが欠如しているデータをアンバランスなパネルデータ（unbalanced panel data）という。たとえば企業が合併したり倒産するとその企業のデータは得られなくなる。そのような企業データを含むのがアンバランスなデータの典型である。企業や家計のパネルデータはアンバランスであることが多い。

（パネルデータの意義）

わが国では企業財務データ、地域経済データが利用可能な代表的なパネルデータである。したがって企業行動や地域経済動向の分析にはパネル分析がよく行われる[1]。このパネルデータを使用

することの利点としてBaltagiやHsiaoは以下のような点を上げている。

①　経済主体間の異質性をコントロールすることができる。たとえば旧財閥系企業と新興企業では借り入れ行動に差があることが指摘されている。クロスセクションデータには旧財閥系企業もあれば新興企業もあるだろう。しかし時系列データでは旧財閥系企業の性質は時点を通じて変わらない（time　invariant）ので、その差をとらえることはできない。

逆に時間を通じては変化するが、ある一時点では共通するというデータもある。たとえば代表的な嗜好品でありかつ税収も多いタバコの販売価格とビールの販売価格は全国共通である。一時点の都道府県別データ（クロスセクションデータ）でタバコとビールの消費需要関数を分析しようとしても、両者の相対価格の効果を知ることはできない。

いずれのケースもパネルデータであればそれらをコントロールすることができる。

②　サンプル数が増え自由度が増す。変数間の変動がより起きて多重共線関係が起こりにくい。

Ｎ個の経済主体のＴ期間のデータを集めればサンプルはNTとなり、時系列データ（Ｔ個）やクロスセクションデータ（Ｎ個）に比べて自由度が増える。その結果検定の信頼性が高まる。

また金利（銀行預金金利と郵貯金利）は同じ時点で同じような動きを示す。時系列データで銀行預金金利と郵貯金利を用いれば多重共線関係が起こるであろう。地域別に見た場合には銀行の金利は異なるので、その変動を合わせて利用するパネルデータでは、この問題を回避できる可能性がある。

③　動学的最適化や動学的調整問題を分析することができる。

たとえば地域別の失業率の変化や地価の変動を分析するには、パネルデータが不可欠である。さらにパネルデータは経済主体固有の性質からくる異質性を除去した後の効果を見る上でも有益である。たとえば日本の地域経済データに関していえば首都東京は独特の地位を占める。その東京の異質性の効果を

$$y_{it}=a+b_1x_{it}+b_2TokyoDummy_i+e_{it} \qquad 10.1)$$

のように、ｉ＝東京の場合１とし、その他の場合０とする０－１のダミー変数（$TokyoDummy_i$）で見ることもできるであろう。しかし差分を取ること（すなわち$\triangle y_{it}=y_{it}-y_{it-1}$）によって

$$\triangle y_{it}=a'+b_1\triangle x_{it}+u_{it} \qquad 10.2)$$

とし、ｘの変動がｙの変動に与える効果を見ることができる。これによって直接観察されないtime invariantな変数の影響を除去できる（もちろん観察されるtime　invariantな変数の影響も除去できる）。

④　より複雑な動きを分析できる。

生産がフロンティア曲線上にあるかどうか、あるいはある企業の倒産確率がどのように変化しているかは個別企業特有の事情とその時々の特殊事情の双方をコントロールする必要がある。これはパネルデータを用いてコントロールすることができる。

（パネルデータの問題）

パネルに特有の問題もある。それはサンプルから落ちていく（attrition）問題である。企業が倒産すればその企業はサンプルから除去され、アンバランスドパネルとなる。このことは家計のパネル分析ではしばしば問題となる。その場合でもア

１）日本では家計のパネルデータは極めて乏しい。国際的に有名なパネルデータとしてミシガン大学のPanel Study of Income Dynamics（PSID）、オハイオ大学のNational Longitudinal Surveys of Labor Market（NLS）がある。

ンバランスドデータにより分析を進めることができる[2)]。

より大きな問題はパネルデータは、通常の例では経済主体の数（N）は多いが、期間（T）は短いということである。様々な仮説の検定は大標本理論に依存している。パネルデータの多くは年単位で作成されるのでTは決して大きくはない。そのためにNを十分多く確保しておく必要がある。またTが十分大きいことを前提とする仮説検定方法や推計方法は使用を避けることが望ましい。

実際にもパネルデータは２時点間で分析されることがある。上場企業の財務分析では10年間もあれば、Tとしては多い方である[3)]。

2　パネルデータのOLS推計

以下説明の便宜のために、データはバランスしているものとする（アンバランスの場合でも本質は変わらない）。

$y_{it}=a+bx_{it}+e_{it}$　　10.3）

を考える。e_{it}〜iid（0，σ^2）であり、a、bが時間を通じ一定かつ経済主体を通じ一定であれば単純にOLS推定を行うことができる（パネル分析との対比でplain OLSということがある）。この時パネルデータの特殊性や経済主体特有の効果は考えられていない。

次に以下のようなモデルを考える。

$y_{it}=a+bx_{it}+e_{it}$　　10.4a）

誤差項について

$e_{it}=\alpha_i+v_{it}$　　10.4b）

という構造を考える。x_{it}はv_{it}と相関しない、v_{it}は標準的線形回帰モデルの仮定を充たす誤差項とする。すなわち

$E(v_{it})=0$，$E(v_{it}^2)=V(v_{it})=\sigma_v^2$，

$E(v_{it}v_{js})=Cov(v_{it},\ v_{js})=0$　for $i\neq j$，$t\neq s$

とする。ここでα_iは経済主体特有の効果（individual effect）と呼ばれているものである。

α_iと説明変数x_iが無相関であれば変量効果モデル（variance component model, error component model）、あるいはランダム・イフェクトモデル（random effect model）という。これに対しα_iとx_{it}が相関していれば固定効果モデル、フィックスド・イフェクトモデル（fixed effect model, covariance model）という。具体的には

$E(\alpha_i x_{it})=Cov(\alpha_i,\ x_{it})=0$　　10.5a）

$E(\alpha_i x_{it})=Cov(\alpha_i,\ x_{it})\neq 0$　　10.5b）

で示される。すなわち10.5a）式を充たせばBalestraとNerloveによって提唱されたランダム・イフェクトモデル、10.5b）式を充たせばフィックスド・イフェクトモデルである。

3　フィックスド・イフェクトモデル

3.1　ダミー変数を用いるケース

まずα_iが説明変数x_iと相関する10.5b）式のケースを見てみよう。説明変数と誤差項が相関を持つので、10.4a）式をOLSで推定するとOLS推定量は不偏性も一致性も持たない（第５章参照）。時点が２期であるとして10.1）式と10.2）式を考えよう。

$y_{it}=a+b_1x_{it}+b_2TokyoDummy_i+e_{it}$

$t=1,\ 2,\ i=1,\ \cdots,\ N$　　10.1）

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

10.1）式の差分を取ることによって

$\triangle y_{it}=b_1\triangle x_{it}+\triangle e_{it}$　　$i=1,\ \cdots,\ N$　　10.2）′

となり、$\triangle e_{it}=\triangle v_{it}$は明白である。差分を取るの

2）企業分析を行おうとするとき、倒産企業や合併企業のデータを除いて企業行動を分析することは望ましくない。それを行えば相対的に健全な企業や合併を行わない企業の分析にとどまり、サバイバル・バイアス（survival bisa）が生じる。

3）家計のアンケート調査では、回答拒否や回答誤差（思い違い）の問題もある。これらについては次章のprobit分析で触れることにする。

で時間を通じて一定のtime-invariantな変数（ここでは定数項、α_iとTokyoDummy$_i$）は除去される。このとき

$$E(\triangle x_{it}\triangle e_{it})=Cov(\triangle x_{it},\ \triangle e_{it})$$
$$=Cov(\triangle x_{it},\ \triangle v_{it})=0 \qquad 10.6)$$

となる。これから10.2）'式をOLSで推定すると固定効果モデルの推定量は一致性を持つことが分かる。

つぎにより一般的に期間がT、経済主体がNというケースを考えよう。

フィックスド・イフェクトモデルは経済主体特有の効果、あるいは各時点に特有の効果があるということを意味している。これらはダミー変数でとらえることができる。たとえば経済主体毎のダミー変数を考え、

$$y_{it}=\alpha_1 Dummy1_i+\alpha_2 Dummy2_i+\cdots$$
$$+\alpha_N DummyN_i+bx_{it}+v_{it} \qquad 10.7)$$

とすることが考えられる。Dummyj$_i$は $i=j$ の時１とし、その他の時０とする０－１ダミー変数である。このようにダミー変数を使う方法をダミー変数最小自乗法（least squares dummy variables model, LSDV）ということがある[4]。

経済主体毎に個別効果があるかどうかの検定はＦ検定で行うことができる。個別効果がなければ

$$y_{it}=\alpha_1+bx_{it}+v_{it} \qquad 10.8)$$

である。帰無仮説は個別効果はないとして、

$$H_0: \alpha_1=\alpha_2=\cdots=\alpha_N \qquad 10.9)$$

である。対立仮説H_1：はH_0ではないである。

10.8）式の残差平方和をRSS_r、10.7）式の残差平方和をRSS_uとし、

$$F=\frac{(RSS_r-RSS_u)/(N-1)}{RSS_u/(NT-(N+1))} \qquad 10.10)$$

を自由度N－1、NT－(N＋1)でF検定を行う（NTはサンプル総数である。アンバランスなデータであればその総数となる）。説明変数がｋ個であれば、

$$F=\frac{(RSS_r-RSS_u)/(N-1)}{RSS_u/(NT-(N+k))} \qquad 10.10)'$$

を検定する。帰無仮説が正しければＦは自由度N－1、NT－(N+k)のＦ分布に従う。

時点特有の時間効果もダミー変数を使い

$$y_{it}=c_1 TimeDummy1_t+c_2 TimeDummy2_t+\cdots$$
$$+c_T TimeDummyT_t+bx_{it}+v_{it} \qquad 10.11)$$

をOLSで推計すればよい。この場合TimeDummyj$_t$は $j=t$ の時１とし、その他の時０とする０－１ダミー変数である。経済主体特有の効果の場合と同様に時間効果があるかどうかをＦ検定で検定することができる。時間効果のない場合$c_1=c_2=\cdots=c_T=0$となる[5]。

さらに個別効果と時間効果を共に考えた

$$y_{it}=\alpha+\alpha_2 Dummy2_i+\cdot+\alpha_N DummyN_i$$
$$+c_2 TimeDummy2_t+\cdot+c_T TimeDummyT_t$$
$$+bx_{it}+v_{it} \qquad 10.12)$$

をOLSで推計することができる[6]。

個別効果、時間効果が共に無いという帰無仮説は10.12）式の残差平方和をRSS_{u2}とし、帰無仮説

４）もちろん10.7）式のかわりに
$$y_{it}=\alpha+\alpha_2 Dummy2_i+\cdots+\alpha_N DummyN_i+bx_{it}+v_{it}$$
をOLSで推定し、$H_0: \alpha_2=\cdots=\alpha_N=0$をＦ検定で検定することもできる。

５）もちろん10.11）式の代わりに
$$y_{it}=c+c_2 TimeDummy2_t+\cdots+c_T TimeDummyT_t+bx_{it}+v_{it}$$
をOLSで推定し、$H_0: c_2=\cdots=c_T=0$をＦ検定で検定することができる。

６）ここでも10.12）式の代わりに
$$y_{it}=\alpha+\alpha_1 Dummy1_i+\alpha_2 Dummy2_i+\cdot+\alpha_N DummyN_i$$
$$+c_1 TimeDummy1_t+c_2 TimeDummy2_t+\cdot+c_T TimeDummyT_t+bx_{it}+v_{it}$$
を推定することが考えられる。しかし$\Sigma_j Dummyj_i=\Sigma_k TimeDummyk_t=1$が成立する。すなわち、完全な多重共線関係が存在する。これを避けるために$\Sigma_i\alpha_i=\Sigma_t c_t=0$の制約がおかれている（第１経済主体のダミーと時間効果の第１期ダミーが除かれている）。

を

H_0：$\alpha_2=\cdots=\alpha_N=0$かつ$c_2=\cdots=c_T=0$

とする。つぎのＦ検定を行うことができる。

$$F=\frac{(RSSr-RSS_{u2})/(N+T-2)}{RSS_{u2}/\{NT-(N+T)\}} \qquad 10.10)''$$

10.12）式は

$$y_{it}=\alpha_i+c_t+bx_{it}+v_{it} \qquad 10.13)$$

においてc_tが観察不能な時点特有の効果（ただし個別の経済主体にとっては共通）を表すものと考えることができる。誤差項にv_{it}以外に個別効果のα_iと時間効果c_tの２つの要素を考慮するので、two factor modelといわれることがある（これに対し10.4a）式と10.4b）式をone factor modelということがある）。

以上の例ではパラメータｂは各経済主体、各時点で共通であると仮定していた。パラメータが経済主体毎に異なるという

$$y_{it}=\alpha+\Sigma_{j=2}\alpha jDummy_{ji}+\Sigma_{s=2}csTimeDummy_{st}+(Dummy2i*b2)x_{it}+\cdots+(DummyNi*bN)x_{it}+bx_{it}+v_{it} \qquad 10.14)$$

を考え、Ｆ検定を行うことができる。ただしこの場合は、推定すべきパラメータが１＋(N－１)＋(T－１)＋(N－１)＋１となり、自由度が著しく減少してしまうという問題がある。

3.2　グループ内変換モデル（within-group estimator model）

$$y_{it}=\alpha_i+bx_{it}+v_{it} \qquad 10.15)$$

において個別経済主体（グループ）毎の平均$y_i=$（１/T）$\Sigma_t y_{it}$、x_i、v_iを考える。10.15）式より、下記のy_iのモデルが得られる。

$$y_i=\alpha_i+b_{xi}+v_i \qquad 10.16)$$

そこで10.15）式から10.16）式を引くと

$$y_{it}-y_i=b(x_{it}-x_i)+(v_{it}-v_i) \qquad 10.17)$$

$$i=1,\ \cdots,\ N、t=1,\ \cdots,\ T$$

を得る。これは10.2）式と同じく経済主体特有の効果は除かれている。平均からの偏差を取った変数について回帰するので、全てのデータを利用した10.17）式のOLS推定量をグループ内推定量（within-group estimator, within estimate）ということがある。

ダミー変数を使ったLSDV（10.7）式のOLSによる推定量と平均からの偏差による回帰分析within estimator（10.17）式のOLSによる推定量のbの推定量は同値となるとともに、推定量の分散やt値も同値となる。10.17）式と比較してLSDVは自由度が減少するように見えるが、10.17）式の実際の自由度は全く同じである。10.17）式の場合同時に推定する係数の方が少いので、10.17）式の方が計算上では便利であるが、固定効果の推定値は直接に得られない[7]。

3.3　関連モデルと推定量

まず次のモデルを考える。

$$y_{it}=\alpha_i+b_ix_{it}+v_{it} \qquad 10.18)$$

v_{it}は標準的線形回帰モデルの仮定を充たすとする。推定量を紹介する前にいくつかの記号を定義する。ここでグループ内平均からの偏差を取り次のように定義する（within groups）。

$$w_{xi}=\Sigma_t(x_{it}-\bar{x}_i)^2 \qquad W_x=\Sigma_i w_{xi}$$

$$w_{xyi}=\Sigma_t(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i) \qquad W_{xy}=\Sigma_i w_{xyi}$$

次に全サンプルの平均を$\bar{y}$、$\bar{x}$とし、その偏差を取り以下のように定義する。

$$T_x=\Sigma_i\Sigma_t(x_{it}-\bar{x})^2$$

$$T_{xy}=\Sigma_i\Sigma_t(x_{it}-\bar{x})(y_{it}-\bar{y})$$

７）経済主体が数千もあるようなパネルデータで固定効果を一々解釈することは実際的ではない。また論文に報告するには余りにも煩瑣である。そのような場合固定効果の報告は省略される。

さらに

$B_x = T_x - W_x$

$B_{xy} = T_{xy} - W_{xy}$

を考える（これをbetween　groupsということがある）[8]。

10.18）式の$y_{it} = \alpha_i + b_i x_{it} + v_{it}$を各経済主体毎に推定した場合のOLS推定量は

$$\hat{b}_i = \frac{W_{xyi}}{W_{xi}} \quad \alpha_i = \bar{y}_i - \hat{b}_i \bar{x}_i \qquad 10.19)$$

である。

全サンプルをプールし、xのパラメータも定数項も共通（すなわち$b = b_1 = \cdots = b_N$，$\alpha = \alpha_1 = \cdots = \alpha_N$）であるとして

$$y_{it} = \alpha + b x_{it} + v_{it} \qquad 10.20)$$

を推計したならば

$$\hat{b} = \frac{T_{xy}}{T_x} = \frac{W_{xy} + B_{xy}}{W_x + B_x} \quad \alpha = \bar{y} - \hat{b}\bar{x}$$

$$\dot{v}_{it} = y_{it} - \alpha - \hat{b} x_{it} \qquad 10.21)$$

である。

全サンプルをプールし、xのパラメータは共通（すなわち$b = b_1 = \cdots = b_N$）であるが、定数項は経済主体毎に異なる

$$y_{it} = \alpha_i + b x_{it} + v_{it}$$

としてフッイクスド・イフェクトモデルを推計したならば

$$\hat{b} = \frac{W_{xy}}{W_x} \quad \alpha_i = \bar{y}_i - \hat{b}\bar{x}_i \quad \tilde{v}_{it} = y_{it} - \alpha_i - \hat{b} x_{it} \qquad 10.22)$$

である。

10.18）式の場合に10.16）式と同じような形にすると

$$\bar{y}_i = \alpha + b\bar{x}_i + \bar{v}_i \quad i = 1, \cdots, N \qquad 10.23)$$

となる。これをOLSで推定すると

$$\hat{b} = \frac{B_{xy}}{B_x} \quad \alpha = \bar{y} - \hat{b}\bar{x} \quad \hat{v}_i = \bar{y}_i - \alpha - \hat{b}\bar{x}_i \qquad 10.24)$$

となる。これをグループ間推定量（between-group estimator, between estimate）ということがある。

4　ランダム・イフェクトモデル

10.4a）式と10.4b）式を再び考えよう。

$$y_{it} = a + b x_{it} + e_{it} \qquad 10.4a)$$

$$e_{it} = \alpha_i + v_{it} \qquad 10.4b)$$

x_{it}はv_{it}と相関しない、v_{it}は標準的線形回帰モデルの仮定を充たす誤差項とする。すなわち

$E(v_{it}) = 0$，$E(v_{it}^2) = V(v_{it}) = \sigma_v^2$,

$E(v_{it}v_{js}) = Cov(v_{it}, v_{js}) = 0$　for $i \neq j$, $t \neq s$

$E(\alpha_i x_{it}) = Cov(\alpha_i, x_{it}) = 0$

の条件が充たされているケースを考えよう。

α_iに関し

$E(\alpha_i) = 0$，$E(\alpha_i^2) = V(\alpha_i) = \sigma_\alpha^2$,

$E(\alpha_i \alpha_j) = Cov(\alpha_i, \alpha_j) = 0$　for $i \neq j$

$E(\alpha_i v_{it}) = 0$,

を仮定する。そうするとe_{it}は

$E(e_{it}) = 0$、

$$E(e_{it}e_{js}) = Cov(e_{it}, e_{js}) = \begin{cases} \sigma_\alpha^2 + \sigma_v^2 & \text{for } i = j,\ t = s \\ \sigma_\alpha^2 & \text{for } i = j,\ t \neq s \\ 0 & \text{その他} \end{cases}$$

である。

説明変数と誤差項は相関しないので10.4a）式のOLS推定量も不偏性と一致性は充たす。ただし、誤差項は相関するのでOLSは有効推定量ではない

8）ここで

$$\Sigma_i \Sigma_t (x_{it} - \bar{x})(y_{it} - \bar{y}) = \Sigma_i \Sigma_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \Sigma_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})$$

を簡単に証明できる。すなわち$B_{xy} = \Sigma_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})$。

9）以下の説明はMaddala［1977］, Kmenta［1997］による。

し、仮説検定も問題になる[9]。BLUEを得る一つの解決策としては、10.4a）式を変換することによって誤差項が相関しなくなるようにすることである。具体的には

$$y_{it}-c\bar{y}_i=a(1-c)+b(x_{it}-c\bar{x}_i)+(v_{it}-c\bar{v}_i) \qquad 10.25)$$

において（１－c)²を次のように

$$\frac{\sigma_v^2}{T\sigma_\alpha^2+\sigma_v^2}$$

選ぶと、$(v_{it}-c\bar{v}_i)$は互いに相関しないことを証明することができる。10.25）式は標準的線形回帰モデルの仮定を充たすので、10.25）式をOLSで推定すればOLSはBLUEになり、仮説検定も行うことができる。これはランダム・イフェクトモデルの一般化最小自乗法の推定量になる。

この推定量は次にのように表現できる。

$$\hat{b}=\frac{W_{xy}+\theta B_{xy}}{W_x+\theta B_x} \qquad \theta=\frac{\sigma_v^2}{T\sigma_\alpha^2+\sigma_v^2} \qquad 10.26)$$

10.26）式と10.21）式を比較すると、10.21）式は10.26）式でθ＝１の特殊ケースであることが分かる。同様に10.22）式は10.26）式でθ＝０の特殊ケースであることが分かる。

T→∞となればθ→０となる。したがって時間が十分多くなればランダム・イフェクトモデルはフイックスド・イフェクトモデルに収斂する。

$\sigma_\alpha^2=0$のケースでは、全サンプルをプールしたOLSの推定量と一致する。

ｃ又はθは未知の係数なので10.25）式をOLSで推定するのは非現実的である。ｃ又はθを推定する必要がある。c、θについては第一段階でフイックスド・イフェクトモデル（LSDV）を推定し、10.22）式の残差（$\tilde{v}_{it}$）より

$$\sigma_v^2=\Sigma\Sigma\tilde{v}_{it}^2/(NT-N-1) \qquad 10.27)$$

とし、これがσ_v^2の推定量となる。10.25）式の残差（$\hat{v}_i$）より

$$\sigma^2=\Sigma\Sigma\hat{v}_i^2/(N-2) \qquad 10.28)$$

とし、これは$\sigma_\alpha^2+\sigma_v^2/T$の推定量となるので$\sigma_\alpha^2$の推定量は

$$\sigma_\alpha^2=\sigma^2-\sigma_v^2/T \qquad 10.29)$$

として得られる。第二段階でこれを用いc又はθを計算する[10]。

（Eviewsによる推定例）

37都県市の地方公営企業（バス事業）の５年分（93―97年度）のデータによりパネル分析を行ってみよう[11]。データが次のように入力されていたとする（図１のパネルAの形式）。_city１，_city２云々は個別経済主体の名前（各地の公共バス事業者）である。経済主体を識別するものである（名前の前または後に_を付ける）。年度（yearと表記する）、生産（seisan，延べ輸送人キロの対数値）、車両数（sharyo，対数値）、労働者数（labor，対数値）が入力されている。仮にファイル名をa：bus.txtとしてAドライブに保存してあるものとする。

_city１	1	13.71478	7.597898	8.258941
_city２	1	12.62907	6.142037	6.408529
_city３	1	12.78309	6.354370	6.981935
………………………				
_city35	1	11.17201	5.293305	5.799093
_city36	1	9.286746	3.295837	3.583519
_city37	1	11.28405	5.214936	5.723585
………………………				
_city１	5	13.67891	7.573017	8.249052
_city２	5	12.47407	6.167517	6.308098
_city３	5	12.55179	6.269096	6.779922
………………………				

10）ランダム・イフェクトモデルに関しても、個別効果と時間効果の双方を考慮したtwo factor model $e_{it}=\alpha_i+c_t+v_{it}$を考えることができる。

11）地方公営企業のデータは横浜市立大学経済学研究科山下耕二氏によって収集された。山下耕二氏に謝意を表します。

_city35	5	11.10607	5.313206	5.799093
_city36	5	9.188197	3.295837	3.583519
_city37	5	11.22961	5.225747	5.843544

このデータを利用して地方公営バス事業者の生産関数

$seisan_{it} = \alpha_i + b_1 sharyo_{it} + b_2 labor_{it} + v_{it}$

の推計を試みよう（ちなみに_city１が東京都_city２が長崎県である）。

Eviewsのパネルのプログラミングは、OLSなどとは若干異なる。まず作業領域を設定する。そのためだけのプログラムを作成する。

workfile a：koutu a 1993 1997

（ａは年次データであることを示す。期間が1993から1997年であることを指定している）

ここで実行し、作業領域を確保する。次にメニューバーで

Objects/Newobject

を選択し実行する。

Type of Objectと聞いてくるので**Pool**を選択する。名前を付けるかどうかを**Name for Object**として聞いてくる。後の作業のために必ずこの名前を付ける。仮にここではpankouと名付けて実行している。

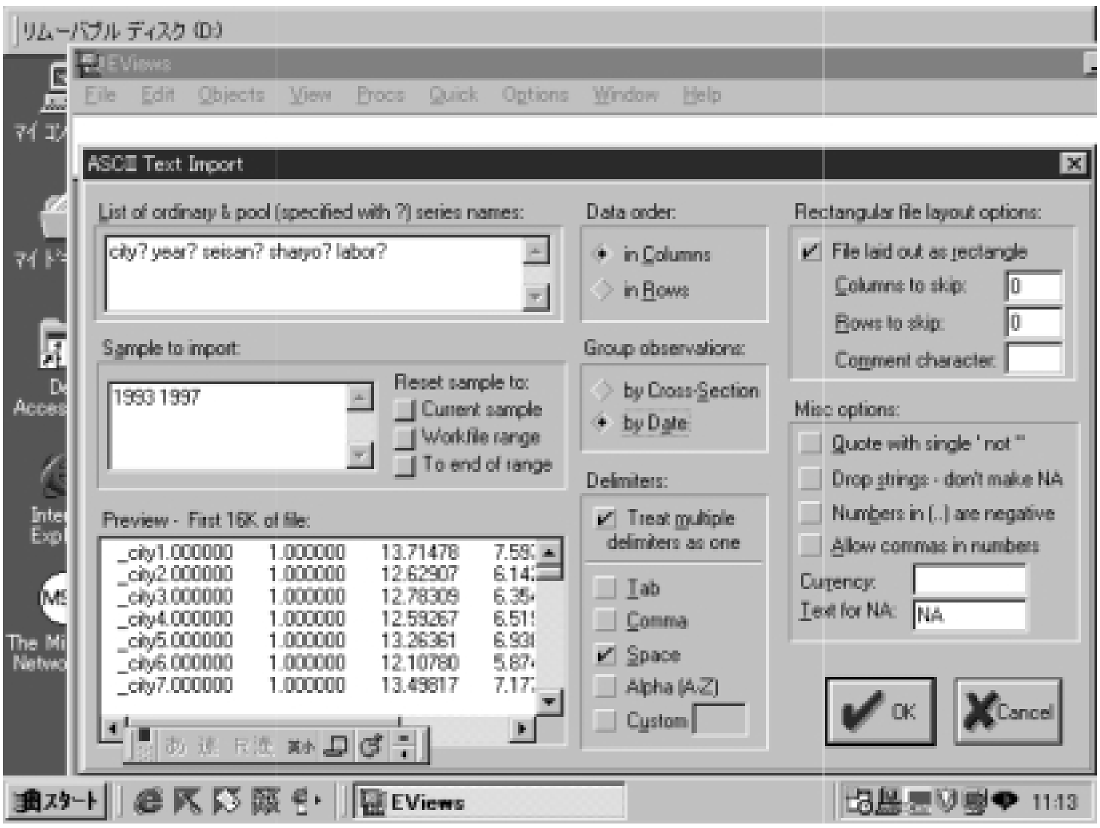
その次に

Cross Section Identifiers:（Enter identifiers below this line）と経済主体の名前（個別主体の識別変数）を聞いてくる。このケースでは_city１から_city37が各々の地方公営バス事業者の名前であるから、これを入力する。**Define**で確定する。

次にデータを読み込む。

Procs/Import Pool data（ASC II,. XLS,. WK？）を選択し実行する。ダイアログ上でa：bus.txtを選択する。

図10.１　パネルデータの読み込み



ASC Ⅱ Text Importの画面となる（図10．1参照）。

List of ordinary & pool［specified with？］series names：
の欄に変数名を入れる。このとき変数の末尾に？を必ず加える（これで個別主体の識別変数と関連付けが行われる）。ここで_**city**は**city**と変更されていることに注意してほしい（_が付けられていない）。

データの入力方法に応じて**Data Order, Group observations, Delimiters**のそれぞれを選択する（ここではパネルAの形式なのでin Columns, **by Date Space**が選択されている）。

次にパネル推計のために**Estimate**を選択する（図10．2参照）。

$seisan_{it}=\alpha_i+b_1 sharyo_{it}+b_2 labor_{it}+v_{it}$を推計するために

Dependent Variableに**seisan？**、**Common coefficients**に**sharyo？ labor？**と入力する（ここでも変数名の後に？を付ける）[12]。

モデルとウエイトの選択を**Intercept**と**Weighting**

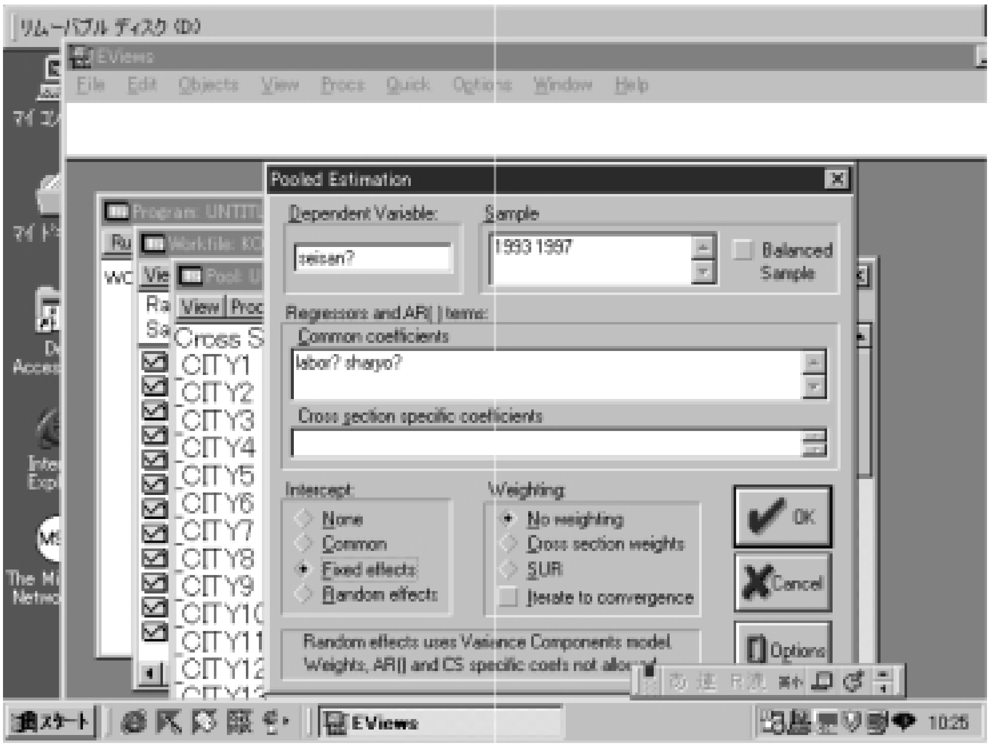
Interceptの**None**は定数項のないモデルである[13]。

Interceptの**Common**とweightingの**No weighting**の組合せはplain OLSに相当する。

Interceptの**Fixed effects**とweightingの**No weighting**の組合せは経済主体毎の効果をダミー変数でとらえたLSDVに相当する。

Interceptの**Common**とweightingの**Cross section weighting**の組合せは、10.20）式においてグループによって誤差項の分散が異なることを考

図10．2　パネル推計



12）Cross section specific coefficientsは経済主体毎にパラメータが異なると予想される場合に用いる。
13）固定効果の標準偏差を求めるときにNoneを選択し、Cross section specific coefficientsの欄に定数項を入れる。

慮する推定方法である。まず10.20）式をplain OLSで推定し、$\dot{\sigma}_i^2=(1/T)\Sigma_{t=1}^{T}v_{it}^2$を計算し、10.20）式を次のように変換する。

$$y_{it}/\dot{\sigma}_i=\alpha/\dot{\sigma}_i+bx_{it}/\dot{\sigma}_i+v_{it}/\dot{\sigma}_i \qquad 10.30)$$

10.30）式をOLSで推定する。これは10.20）式をGLSで推定することと同じである。

Interceptの**Fixed effects**とweightingの**Cross section weighting**の組合せは、10.7）式において、グループによって誤差項の分散が異なることを10.30）式と同様に考慮する推定方法である。

Interceptの**Random effects**とweightingの**No weighting**の組合せは、10.25）式に相当する。

最初にInterceptのFixed effectsとweightingのNo weightingの組合せを選んでLSDVを推計してみよう。結果は表10.１に掲げるとおりである。上欄に推計方法がPooled Least Squares（LSDV）であることが示されている。クロスセクション（経済主体）の数が37、時点が５でサンプル総数が185でバランスしたパネルデータであることが表示されている。

車両（sharyo）の符号は負（統計的には非有意）と、理論条件を充たしていない。労働者数は1.1と１％水準で有意に正である。各バス事業者の固定効果が次に示される（_cityj-cとして表示されている）。東京は6.9（_city１-cの係数）と相対的に高いが、長崎県の7.5（_city２-cの係数）を下回っている。長崎県の方が生産性が高いことが分かる[14)]。

このダミー変数によるLSDVで車両数が延べ人キロで見た生産に影響しない（符号が負）というのは若干不自然である。各経済主体のウエイトが等しいという仮定に問題があるのかもしれない。

次にInterceptの**Fixed effects**とweightingの

表10.１　フィックスド・イフェクトモデル（LSDV）の例

Dependent Variable: SEISAN？
Method: Pooled Least Squares
Sample：1993 1997
Included observations：5
Number of cross-sections used：37
Total panel (balanced) observations：185

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SHARYO？	−0.311907	0.236010	−1.321584	0.1884
LABOR？	1.108237	0.184843	5.995555	0.0000
Fixed Effects				
_CITY１−C	6.914736			
_CITY２−C	7.458211			
_CITY３−C	7.003355			
_CITY４−C	6.810248			
_CITY５−C	6.823765			
_CITY６−C	6.401130			
_CITY７−C	6.902206			
_CITY８−C	7.101887			
_CITY９−C	6.742702			
_CITY10−C	6.879923			
_CITY11−C	6.717834			
_CITY12−C	6.299918			
_CITY13−C	6.925315			
_CITY14−C	6.637848			
_CITY15−C	6.645835			
_CITY16−C	6.213976			
_CITY17−C	6.351174			
_CITY18−C	6.622593			
_CITY19−C	6.700238			
_CITY20−C	6.322370			
_CITY21−C	6.158402			
_CITY22−C	6.278078			
_CITY23−C	6.547486			
_CITY24−C	6.784346			
_CITY25−C	6.070517			
_CITY26−C	6.573282			
_CITY27−C	6.679055			
_CITY28−C	6.277056			
_CITY29−C	6.437909			
_CITY30−C	6.690027			
_CITY31−C	6.435664			
_CITY32−C	6.506064			
_CITY33−C	6.870160			
_CITY34−C	6.470474			
_CITY35−C	6.371661			
_CITY36−C	6.272202			
_CITY37−C	6.493821			

R-squared	0.994421	Mean dependent var	11.29229
Adjusted R-squared	0.992969	S.D. dependent var	1.233076
S.E. of regression	0.103395	Sum squared resid	1.560807
F-statistic	26023.84	Durbin-Watson stat	1.276162
Prob (F-statistic)	0.000000		

14）経済主体の数が数千、数百の時はこの固定効果の報告は省略されるが、本件のように数十という例では報告されることが多い。特に地域間の比較に関心がある場合は報告することが望ましい。

Cross section weightingの組合せによるGLSを行ってみよう（表10．2参照）。

上欄に推計方法がGLS（Cross Section Weights）であることが示されている。車両は0．24と符号は正である。ｔ値は1．78でありｐ値は0．078であるから10％水準で有意である。車両が１％増加すると延べ人キロは0．24％増えることがうかがわれる。労働者数の係数も0．87と正であり、１％水準で有意である。

ウエイト付けしたGLS推計による東京都の固定効果を見ると4．68である。長崎県はもとより他の全ての事業者を下回っている。東京にはバス事業の運営を阻害する特殊要因（たとえば車両が極端に多いことによる混雑）が存在することがうかがわれる。

下欄にウエイト付けしたR^2等が報告されている（論文ではこのウエイト付けしたものを報告する）。

記述統計は**View/Descriptive Statistics**を選択すると

Stacked Data（全サンプルの単純平均）、**Stacked–means removed**（平均からの偏差を取った統計）、**Cross section specific**（経済主体毎の統計）、**Time period specific**（時点毎の統計）がでてくる。ここから必要な統計量を選択すればよい。

残差についても**View/Residuals**で知ることができる。**Graph**を選択すると各経済主体毎の残差が示される（図10．３参照）

次に$b_1+b_2=1$の一次同次の制約が有効かどうかをWald検定してみよう。

View/Wald Coefficient Testsを選択し

C(1)＋C(2)＝１とする（固定効果モデルで定数項が無いのでsharyoは最初の説明変数となる）。

結果はχ^2統計量の値は1．27（p値は0．26）であった。このケースでは一次同時の制約が有効で

表10．2

Dependent Variable: SEISAN？
Method: GLS (Cross Section Weights)
Sample: 1993 1997
Included observations: 5
Number of cross–sections used: 37
Total panel (balanced) observations: 185

Variable	Coefficient	Std. Error	t–Statistic	Prob.
SHARYO？	0．241147	0．135852	1．775068	0．0780
LABOR？	0．870530	0．077562	11．22370	0．0000
Fixed Effects				
_CITY 1－C	4．682762			
_CITY 2－C	5．564057			
_CITY 3－C	5．161005			
_CITY 4－C	4．898389			
_CITY 5－C	4．822407			
_CITY 6－C	4．760987			
_CITY 7－C	4．808985			
_CITY 8－C	5．138368			
_CITY 9－C	4．765791			
_CITY10－C	5．001910			
_CITY11－C	5．308194			
_CITY12－C	4．961924			
_CITY13－C	5．580708			
_CITY14－C	5．014354			
_CITY15－C	5．139471			
_CITY16－C	4．722552			
_CITY17－C	5．045718			
_CITY18－C	5．191697			
_CITY19－C	5．409940			
_CITY20－C	4．893640			
_CITY21－C	4．998920			
_CITY22－C	5．042605			
_CITY23－C	5．354037			
_CITY24－C	5．190062			
_CITY25－C	5．121369			
_CITY26－C	5．423508			
_CITY27－C	5．310194			
_CITY28－C	5．258644			
_CITY29－C	5．208341			
_CITY30－C	5．291622			
_CITY31－C	5．396658			
_CITY32－C	5．557689			
_CITY33－C	5．527065			
_CITY34－C	5．020095			
_CITY35－C	4．811962			
_CITY36－C	5．299917			
_CITY37－C	4．980755			

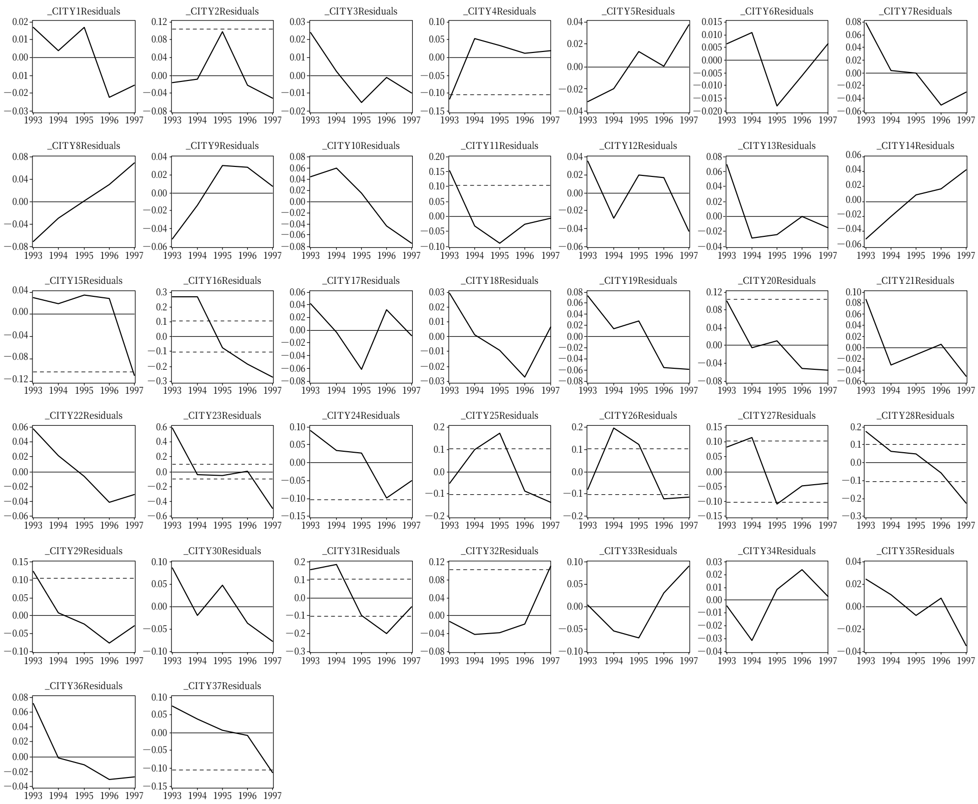
Weighted Statistics

R–squared	0．999986	Mean dependent var	28．21339
Adjusted R–squared	0．999982	S.D. dependent var	22．95388
S.E. of regression	0．097952	Sum squared resid	1．400799
F–statistic	10104180	Durbin–Watson stat	1．452818
Prob (F–statistic)	0．000000		

Unweighted Statistics

R–squared	0．994210	Mean dependent var	11．29229
Adjusted R–squared	0．992703	S.D. dependent var	1．233076
S.E. of regression	0．105335	Sum squared resid	1．619938
Durbin–Watson stat	1．194140		

図10．3



ある。

InterceptのRandom　effectsとweightingのNo weightingの組合せによりランダム・イフェクトモデルを推計してみよう（表は10．３参照）。

上欄にGLS（Variance Components）として推計方法がランダム・イフェクトモデルであることが示される。車両は符号は正であるものの統計的には有意ではない。労働者数の値は0.88であり、１％水準で有意である。この結果はLSDVによったフィックスド・イフェクトモデルと共通している。

ランダム・イフェクトモデルでも個別の定数項（Random　Effects）をEviewsは表示する。これは以下の10.31）式によって求められたものである。

$$a_i=\frac{\sigma_\alpha^2}{T\sigma^2}\Sigma_t(y_{it}-\hat{a}-\hat{b}_1 sharyo_{it}-\hat{b}_2 labor_{it}) \quad 10.31)$$

ここでσ$_\alpha^2$とσ^{2}はそれぞれ10.29）式と10.28）式のように計算される。

_city16（ちなみにこれは秋田市である）が−0.38と最も大きな負の効果が見られる。

５　フッイクスド・イフェクトモデルとランダム・イフェクトモデルの選択

（選択の考え方）

E($\alpha_i x_i$)＝Cov($\alpha_i x_i$)＝０であればランダム・イフェクトモデル、E($\alpha_i x_i$)＝Cov($\alpha_i x_i$)≠０であればフッイクスド・イフェクトモデルが選択されるとした。

これは経済主体（企業など）の観察不可能な固有の要素が観察可能な説明変数と相関しないのか、あるいは相関するするのかということでもある。同一産業内には多くの企業が存在する。その業績

表10.3　ランダム・イフェクトモデルの推計

Dependent Variable: SEISAN？
Method: GLS （Variance Components）
Sample：1993　1997
Included observations：5
Number of cross-sections used：37
Total panel（balanced）observations：185

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.803595	0.156323	37.12569	0.0000
SHARYO？	0.097145	0.143963	0.674795	0.5007
LABOR？	0.880837	0.130637	6.742643	0.0000
Random Effects				
_CITY1−C	−0.107717			
_CITY2−C	0.550853			
_CITY3−C	0.182341			
_CITY4−C	−0.039943			
_CITY5−C	−0.056970			
_CITY6−C	−0.253016			
_CITY7−C	−0.040189			
_CITY8−C	0.222956			
_CITY9−C	−0.120071			
_CITY10−C	0.056571			
_CITY11−C	0.152880			
_CITY12−C	−0.216390			
_CITY13−C	0.377580			
_CITY14−C	−0.064262			
_CITY15−C	0.017432			
_CITY16−C	−0.380403			
_CITY17−C	−0.134208			
_CITY18−C	0.056352			
_CITY19−C	0.223553			
_CITY20−C	−0.223183			
_CITY21−C	−0.230977			
_CITY22−C	−0.158259			
_CITY23−C	0.071523			
_CITY24−C	0.115133			
_CITY25−C	−0.246665			
_CITY26−C	0.144994			
_CITY27−C	0.098790			
_CITY28−C	−0.070986			
_CITY29−C	−0.026292			
_CITY30−C	0.113361			
_CITY31−C	0.069601			
_CITY32−C	0.184296			
_CITY33−C	0.302325			
_CITY34−C	−0.112544			
_CITY25−C	−0.272055			
_CITY36−C	−0.062563			
_CITY37−C	−0.123848			

GLS Transformed Regression

R-squared	0.992857	Mean dependent var	11.29229
Adjusted R-squared	0.992779	S.D. dependent var	1.233076
S.E. of regression	0.104783	Sum squared resid	1.998275
Durbin-Watson stat	0.966439		

Unweighted Statistics including Random Effects

R-squared	0.994224	Mean dependent var	11.29229
Adjusted R-squared	0.994161	S.D. dependent var	1.233076
S.E. of regression	0.094227	Sum squared resid	1.615934
Durbin-Watson stat	1.195105		

は様々である。生産性の高い企業もあれば低い企業もある。その産業の生産関数の推定を試みるとする。外部から観察不可能な要素（たとえば経営者の資質や企業風土）が設備や資金という投入要素と関連しなければランダム・イフェクトモデルを選択すればよい。逆に経営者の資質や企業風土が投入要素に影響すると考えるのならばフッイクスド・イフェクトモデルを選択すればよい。

（Wu-Hausmanテスト）

これを統計的に検定しようというのが第７章で取り上げたWu-Hausmanテストである。

$Cov(\alpha_i X_i)=0$であればランダム・イフェクトモデルは一致性をもちかつ有効性もある。これに対しフイックスド・イフェクトモデルは一致性は有するが有効性はない。

$Cov(\alpha_i X_i)\neq 0$であればフイックスド・イフェクトモデルは一致性を持ちかつ有効推定量である。しかしランダム・イフェクトモデルは一致性も有しない。そこで

帰無仮説H_0：$Cov(\alpha_i X_i)=0$の下での推定量を$\hat{b}_{RE}$とし、

対立仮説H_1：$Cov(\alpha_i X_i)\neq 0$の下での推定量を$\hat{b}_{FE}$とする。

両者の差$\hat{q}=\hat{b}_{FE}-\hat{b}_{RE}$は確率極限において

H_0：$\hat{q}=0$

H_1：$\hat{q}\neq 0$

となる。$\hat{q}$の分散$V(\hat{q})=V(\hat{b}_{FE})-V(\hat{b}_{RE})$を考え

$$m=\frac{\hat{q}^2}{V(\hat{q})} \qquad 10.32)$$

を検定する。この統計量が帰無仮説の下で自由度k（説明変数の数）のχ^2分布に従う[15)]。

（補助回帰による検定）

10.25）式をベースにして上記のHausmanテストをある回帰モデルから得られる10.25）式にあるcの推定量を$\hat{c}$とし、

$\dot{y}_{it} = y_{it} - c\,\bar{y}_i$

$\dot{x}_{it} = x_{it} - c\,\bar{x}_i$

とする。固定効果モデルの10. 17）式の平均からの偏差を取る形の

$\tilde{x}_{it} = x_{it} - \bar{x}_i$

とする。次の式を推計する。

$\dot{y}_{it} = a + \dot{b}x_{it} + d\tilde{x}_{it} + \text{残差}$　　10. 33）

ランダム・イフェクトモデルが正しければ新たに含められた$\tilde{x}_{it}$は説明力を持たないであろう。そこでｄ＝０のｔ検定（$\tilde{x}_{it}$が複数であればＦ検定）を行う。ｄ＝０が棄却できなければランダム・イフェクトモデルを採択する。

（EviewsによるWu-Hausman検定）

次にWu-Hausmanテストを行ってみよう。

先の推定はD：koutaに保存されているとする。またパネル分析の結果はkoutaの中にkoupanと名付けて保存されているものとする。

’作業領域を設定する

workfile a：haus u　1 1000

’Dドライブのkoutaを読み込む。

load d：koutu

’全てのサンプルを用いることを宣言する

smpl @all

’フイックスド・イフェクトを推計し、結果を保存する。

’Fixed effectsとweightingのCross section weightingの組合せ（f、w）がオプションである。仮にウエイト無しを選択するのであれば（f）と変更すればよい。

koupan. ls（f, w）seisan？ sharyo？ labor？

’パラメータをbetaと名付けて保存[16)]

vector beta＝koupan.@coefs

’分散共分散をcovarと名付けて保存

matrix covar＝koupan.@cov

’説明変数のパラメータのみを保存（b_fixed）

vector b_fixed＝@subextract(beta,1,1,2,1)

’説明変数の分散共分散のみを保存（cov_fixed）

matrix cov_fixed

＝@subextract(covar,1,1,2,2)

’ランダムイフェクトモデルを推計

koupan. ls（r）seisan？ sharyo？ labor？

’パラメータをbetaと名付けて保存[17)]

beta＝koupan.@coefs

’分散共分散をcovarと名付けて保存

covar＝koupan.@cov

’説明変数のパラメータのみを保存（b_gls）[18)]

vector b_gls＝@subextract(beta,2,1,3,1)

’説明変数の分散共分散のみを保存（cov_gls）

matrix cov_gls

＝@subextract(covar,2,2,3,3)

’Hausmanテストの計算

’$q = b_{FE} - b_{RE}$を計算

matrix b_diff＝b_fixed－b_gls

’qの分散$V(q) = V(b_{FE}) - V(b_{RE})$を計算

matrix var_diff＝cov_fixed－cov_gls

$m = \dfrac{q^2}{V(q)}$を計算[19)]

matrix qform

＝@transpose(b_diff)*@inverse(var_diff)*b__diff

’qformは非負のはず。その条件を充たす場合結

15）説明変数がkの場合mを下記のように表現する。

$m = q'[V(q)]^{-1}q$

ここでqはベクトルで、$V(q)$は行列になる。

16）betaに保存されるのは**sharyo？**と**labor？**の推定値のみである。

17）フイックスド・イフェクトと違って’ランダム・イフェクトを推定するとbetaに保存されるのは定数項と**sharyo？**と**labor？**の推定値のみである。

18）前注で説明したようにbetaに定数項が含まれ、それに対応するものも共分散行列（covar）にも含まれるのでsubextractを利用し、この分を取り除く。

果を報告する

```
  if qform (1,1) >= 0 then
    '表の作成
    table (4,2) result
    setcolwidth (result,1,15)
    setcell (result,1,1, "Hausman test ")
    setcell (result,2,1, " (fixed versus random effects) ")
    setline (result,3)
    ! df=@rows (b_diff)
    setcell (result,4,1, " Chi–square (" +@str(! df)+" d.f.) "," r ")
    setcell (result, 4, 2, qform (1,1))
    setcell (result, 5, 1, "p–value "," r ")
    setcell (result, 5, 2, 1-@cchisq (qform(1,1), ! df))
    show result
'qformが負となる場合、Quadratic form is negativeと報告する
  else
    statusline" Quadratic form is negative "
  endif
```

結果は表10.4に示すとおりである。

χ^2統計量は2.015である。p値は0.365である。帰無仮説は棄却されず、このケースではランダム・イフェクトモデルが採択される。

表10.4　Hausmanテスト

Hausman test (fixed versus random effects)	
Chi–square (2 d.f.)	2.0150174
p–value	0.3651275

参考文献

kmenta. J［1997］Elements of Econometics（second ech）, Michiganのch12–2

Johnston. D and J. DiNardo［1997］Econometric mechods（4theds）, McGraw Hill, のch12

Maddala. G［1977］Econometrics, McGraw Hill, のch14

19）注15で説明したようにmを正しく表現するために行列表現を使用する必要がある。