パネルデータの分析†

担当: 長倉大輔

† この資料は私のゼミにおいて使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。

パネルデータの分析

- パネルデータとは？

時系列データ:

時間の経過とともに観測されるデータ

(例) 株価、GDP、為替

横断面(クロスセクション)データ:

多数の個体（企業、家計、国など）のある1時点において観測されるデータ

(ある時点の)テストの点数、身長、体重、等

パネルデータの分析

パネルデータ:

上記の2つのデータの特徴を併せ持つデータ。典型的なものは多数の個人について連続した複数の時点で観測されたデータ。

個別の企業や家計のパネルデータを**ミクロパネルデータ**、都道府県や国ごとに集計されたパネルデータを**マクロパネルデータ**という。

パネルデータの分析

- パネルデータの表し方

パネルデータは個体(i)と時点(t)の両方に依存しているので、表し方としてこの2つを下付き文字として

$$Y_{i,t}\ ,\ i=1,\ldots,N,\ \ t=1,\ldots,T$$

のように表す。ここでNは個体数、Tは観測時点数を表す。例えばY_{35}は3番目の個体の第5時点におけるデータを表す。

パネルデータの分析

パネルデータの例:

年度(T 年)

個人 (N人)

$i \setminus t$	1	2	3	…	T
1	Y_{11}	Y_{12}	Y_{13}	…	Y_{1T}
2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2T}
3	Y_{31}	Y_{32}	Y_{33}	…	Y_{3T}
:	:	:	:	:	:
N	Y_{N1}	Y_{N2}	Y_{N3}	…	Y_{NT}

パネルデータの分析

パネルデータは典型的には N(個体数)が大きくて、T (観測時点数)が少ない場合が多い。

家計にアンケートをとってとる調査であれば、T=5, N=200 のようなデータになる場合が多い。

これは T を増やすことに比べて、Nを増やすことが比較的容易なのでこのようになる。

パネルデータの分析

- パネルデータを用いる利点

(1) 個別経済主体間の異質性をコントロールするのが容易である。

(2) 経済主体を個別に分析する場合、T が小さく、精度がよい分析ができない場合でも、パネルデータを用いれば、NT 個のデータを用いることができ、分析の精度が上がる。

などがある。特に(1)が重要。

パネルデータの分析

- パネルデータのモデル

パネルデータ分析には次の形の回帰式がよく用いられる。

$$Y_{it} = \boldsymbol{\alpha_i} + \beta_1 X_{1it} + \ldots + \beta_K X_{Kit} + \varepsilon_{it}$$

ここで

Y_{it}: 個体(i)と時点(t)に依存した被説明変数。

$X_{k,it}$: 個体(i)と時点(t)に依存する説明変数。

α_i: 個体(i)に依存して**異なるが、時間を通じては一定**の切片。

ε_{it}: $E(\varepsilon_{it}) = 0,\ \mathrm{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$ の誤差項で、すべての i, t について独立とする。

パネルデータの分析

- 個別効果

このモデルの一番の特徴は **α_i が個体毎に異なる**という事である $(\alpha_j \neq \alpha_k,\ j \neq k)$。

これによって個体間の異質性をコントロールしている。

この α_i は**個別効果**と呼ばれる。

パネルデータの分析

個別効果は個体間の異質性を切片の違いとしてとらえたものだが、その経済的な解釈は分析者が行う。

例)　企業間の経営能力の差を表すなど。

パネルデータの分析

- パネルデータ分析の注意点

先程のモデルで $K=1$ として Z_i を**個体ごとに異なるが、時間を通じて一定**の変数としよう。

この変数を説明変数として加えると

$$Y_{it} = \alpha_i + \gamma Z_i + \beta X_{it} + \varepsilon_{it}$$

となる。ここで $\delta_i = \alpha_i + \gamma Z_i$ とおけば

パネルデータの分析

$$Y_{it} = \delta_i + \beta X_{it} + \varepsilon_{it}$$

となる。ここで δ_i は時間を通じて一定である。

このように α_i と γZ_i は δ_i という一つの値にまとめられ、分けて考えることはできない。

これは γ は推定できない事を意味する。

パネルデータの分析

つまり、パネルデータにおいて

時間を通じて一定の変数は説明変数 として加えることはできない

という事である。

（より厳密には、個別効果を加えなければ、個体間で異なるが、時間を通じて一定の変数を1つだけ加えることができる。ただし2つ以上は加えられない)。

パネルデータの分析

個別効果に加えて、個体間で異なるが、時間を通じて一定の変数を加える、もしくは個別効果を加えないで、時間を通じて一定の変数を 2つ以上モデルに入れると、完全な多重共線性の問題が生じ、推定できない。

例) 家計のデータでは性別ダミーなどを入れることができない。

パネルデータの分析

- 時間効果

先程のモデルにはさらに、個体間で同じであるが時間を通じて異なる切片を加えることができる。

$$Y_{it} = \alpha_i + \boldsymbol{\theta_t} + \beta_1 X_{1it} + \ldots + \beta_K X_{Kit} + \varepsilon_{it}$$

これは個体間に共通の影響を与える変数の影響を全てあわせたものである。
(例) 家計のデータであれば景気の影響、など。

θ_t は **時間効果** と呼ばれる。

パネルデータの分析

以後は説明の簡単化のため、時間効果はないとする。

パネルデータの分析

- Pooled OLS (集計最小二乗法)

先程のモデルにおいてすべての個体において個別効果が同じである、すなわち

$$\alpha_1 = \alpha_2 = \ \ldots \ = \alpha_N$$

である場合(つまり**個別効果がない**という事)、この共通の α と説明変数の係数 β_k は 通常の最小二乗推定によって推定できる。すなわち

パネルデータの分析

$$Y_{it} = \alpha + \beta_1 X_{1it} + \ \dots + \beta_K X_{Kit} + \varepsilon_{it}$$
$$i = 1, \dots, N, \ \ t = 1, .., T$$

を最小二乗法で推定するだけである。

これはただのOLSによる推定だが、パネルデータにおいて個別効果がないという仮定の下で用いた事を強調するため **Pooled OLS** と呼ばれる。

パネルデータの分析

- 固定効果モデル

これに対して、個別効果が存在する場合、個別効果も未知パラメーターとして推定した場合、**固定効果モデルを推定する**という。

これはモデルというよりは個別効果を推定するかどうかの事を指しているが、慣例として固定効果モデルと呼ぶ。

パネルデータの分析

推定法としては以下の**最小二乗ダミー変数推定**
(Least square dummy variable estimation)
がよく用いられる。

パネルデータの分析

- Least Square Dummy Variable 推定 (最小二乗ダミー変数推定)

説明の簡単化のために以後は説明変数は一つ (K = 1) とする。

第 i 番目の個体に対して、D_{jt}, j =1,…,N を $j = i$ であれば 1をとり、$j \neq i$ であれば 0 をとるダミー変数とする。

パネルデータの分析

このようなダミー変数を用いると、個別効果を含んだモデル

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

は

$$Y_{it} = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_N D_{Nt} + \beta X_{it} + \varepsilon_{it}$$

と書き表すことができる。

個別効果は(説明変数) D_{jt} の係数になっている。

パネルデータの分析

例えば、$i = 2$ の時

$$Y_{2t} = \underset{(=0)}{\alpha_1 D_{1t}} + \underset{(=1)}{\alpha_2 D_{2t}} + \ldots + \underset{(=0)}{\alpha_N D_{Nt}} + \beta X_{2t} + u_{2t}$$

$\Leftrightarrow$

$$Y_{2t} = \alpha_2 + \beta X_{2t} + u_{2t}$$

である。

パネルデータの分析

このように $Y_{it} = \alpha_i + \beta X_{it} + u_{it}$ を

$$Y_{it} = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_N D_{Nt} + \beta X_{it} + \varepsilon_{it}$$

と書き直してあげれば、このモデルにおいて、個別効果 $\alpha_1, \ldots, \alpha_N, \beta$ は $D_{1t}, \ldots, D_{NT}, X_{it}$ を説明変数とした線形回帰モデルの係数であるので最小二乗法によって推定できる。

このような推定方法を**最小二乗ダミー変数推定**と呼ぶ(または***Within*** **推定**とも呼ばれる) 。

パネルデータの分析

- 固別効果の検定

もし固別効果が存在するのであれば、最小二乗ダミー変数推定が望ましい。

他方で、もし個別効果がないのであれば Pooled OLS による推定の方が効率的である事が知られている。

個別効果があるかどうかは F 検定によって検定できる(R でのやり方は後ほど)

パネルデータの分析

- 変量効果モデル

個別効果を含んだモデルにおいて、個別効果を未知パラメーターとして推定するのではなく、確率変数とみなして、（その平均、分散と共に）β を推定する事を**変量効果モデルを推定する**という。

これもモデルというよりは推定法の違い（および推定する対象の違い）といった方がよいが、慣例で変量効果モデル(を推定する）という。

パネルデータの分析

変量効果モデルは、α_i を確率変数とみなし、以下のような想定をする

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}, \ \ E(\alpha_i) = \mu_\alpha, \ \ \mathrm{var}(\alpha_i) = \sigma_\alpha{}^2,$$

α_i と α_j $(i \neq j)$ は独立、α_j と ε_{it} は独立。

この時、$\beta, \mu_\alpha, \sigma_\alpha{}^2$ は**一般化最小二乗法**によって推定できる(より正確にいうと β を一般化最小二乗法によって推定でき、μ_αと $\sigma_\alpha{}^2$はそれに付随した方法で推定する)。

パネルデータの分析

- 変量効果モデルの注意点

変量効果モデルによって β を推定する場合、推定量が一致性を持つためには　個別効果 α_i と説明変数 X_{it} に**相関がない**という条件が必要である($\mathrm{cov}\,(X_{it}, \alpha_i) = 0$)。

もし個別効果と説明変数に相関がある場合は変量効果モデルによる推定は一致性を持たずバイアスが存在する(内生性の問題)。

パネルデータの分析

他方、固定効果モデルでは α_i も未知パラメーターとして推定するため、例え個別効果と説明変数に相関がある場合でも β を一致推定できる。

このように考えると固定効果モデルの方がよいように思えるが、もし**個別効果と説明変数に相関がないという条件が満たされる**のであれば、実は変量効果モデルの方が β の推定効率がよい。

パネルデータの分析

まとめると、

ケース1: 個別効果と説明変数間に相関がない
→ 変量効果モデルによる推定の方がよい
(固定効果モデルは推定効率が落ちる。)

ケース2: 個別効果と説明変数間に相関がある
→ 固定効果モデルによる推定の方がよい。
(変量効果モデルでは一致推定できない)

という事になる。

パネルデータの分析

ただし、分析によっては個別効果の推定により興味がある場合があるのでその場合は固定効果モデルの方がよいだろう。

これらの事より、教科書や論文によっては

$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$ において、

α_i と X_{it} に**相関がある**モデルを**固定効果モデル**

α_i と X_{it} に**相関がない**モデルを**変量効果モデル**

と呼ぶものがある。

パネルデータの分析

- ハウスマン検定

個別効果と説明変数に相関があるかどうかを検定する方法として**ハウスマン検定**がある。

ハウスマン検定によって、帰無仮説

H_0: 個別効果と説明変数に相関がない

が棄却されれば固定効果モデル、棄却されなければ変量効果モデルを用いる。

R によるやり方は後ほど。

パネルデータの分析

- パネルデータの分析の流れ

まず

(1) F 検定によって帰無仮説「個別効果なし」を検定

棄却されない → Pooled OLS

棄却される → (2) へ

パネルデータの分析

(2) ハウスマン検定によって帰無仮説
「個別効果と説明変数に相関なし」
を検定

棄却されない → 変量効果モデルの推定
棄却される → 固定効果モデルの推定

となる。

パネルデータの分析

- 一般化最小二乗法によるパネルデータ回帰式の推定

確率変数 μ_i は $E(\mu_i) = 0,\ \mathrm{var}(\mu_i) = \sigma_\alpha{}^2$ および μ_i と $\mu_j\ (i \neq j)$ は独立、である確率変数とする。

この時、α_i は

$$\alpha_i = \mu_\alpha + \mu_i$$

と表す事ができる。

パネルデータの分析

これを先ほどの式に代入すると

$$Y_{it} = \mu_{\alpha} + \beta X_{it} + u_{it} \quad ,$$

と表すことができる。ここで $u_{it} = \mu_i + \varepsilon_{it}$ である。

パネルデータの分析

ここで、もともとのモデルにおける仮定:

(A1) $E(\mu_i) = 0$, $\mathrm{var}(\mu_i) = \sigma_\alpha{}^2$, μ_i, μ_j $(i \neq j)$ は独立

(A2) $E(\varepsilon_{it}) = 0$, $\mathrm{var}(\varepsilon_{it}) = \sigma_\varepsilon{}^2$,
$\varepsilon_{it}, \varepsilon_{js}$ $(i \neq j$ もしくは $t \neq s)$ は独立

に加え、

(A3) μ_i と ε_{jt} が全ての i, j, t について独立

であるとすると、u_{it} の期待値、分散、共分散は

パネルデータの分析

$$E(u_{it}) = 0, \quad \mathrm{var}(u_{it}) = \sigma_\alpha{}^2 + \sigma_\varepsilon{}^2$$

$$\mathrm{cov}(u_{it}, u_{js}) = \begin{cases} 0, & i \neq j, \\ \sigma_\alpha^2, & t \neq s \ (\text{かつ}\, i = j) \end{cases}$$

となる(演習問題1参照)。

パネルデータの分析

よって、結局パネルデータ回帰式は

$$Y_{it} = \mu_\alpha + \beta X_{it} + u_{it} \quad ,$$

$$E(u_{it}) = 0, \quad \mathrm{var}(u_{it}) = \sigma_\alpha{}^2 + \sigma_\varepsilon{}^2$$

$$\mathrm{cov}(u_{it}, u_{js}) = \begin{cases} 0, & i \neq j, \\ \sigma_\alpha^2, & t \neq s \ (\text{かつ} \ i = j) \end{cases}$$

となる。これは誤差項が不均一分散かつ互いに相関がある回帰モデルであり、以下に述べる一般化最小二乗法で推定できる。

パネルデータの分析

- 一般化最小二乗法

一般化最小二乗法について説明する。
次の(行列で表現した) 回帰式を考えよう。

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

ここで $\mathbf{Y}$ は $N\times1$ベクトル、$\mathbf{X}$ は $N\times K$ ベクトル、$\boldsymbol{\beta}$ は $K\times1$ ベクトル、$\boldsymbol{\varepsilon}$ は $N\times1$ベクトルである。

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ X_{21} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots \\ X_{N1} & \cdots & X_{NK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

パネルデータの分析

- 一般化最小二乗法

通常、最小二乗法により推定する場合は誤差項 $\varepsilon_1, \ldots,$ ε_N は互いに無相関($\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$)で、均一な分散 ($\mathrm{var}(\varepsilon_i) = \sigma^2$) を持つと仮定される。

この時、$\boldsymbol{\varepsilon}$ の分散共分散行列 $\boldsymbol{\Sigma} = E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}^{\mathrm{T}}]$は

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_N$$

となる。ここで $\mathbf{I}_N$ は $N \times N$の単位行列である。

パネルデータの分析

しかしながら実際のデータはこの仮定を満たさない場合も多い。このような時、最小二乗法は**一致性は持つ** (!)が、一般化最小二乗法の方が漸近的には**より効率的**である事が知られている。一般化最小二乗法は誤差項の分散共分散行列として、一般的な形

$$\mathbf{\Sigma} = \mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{bmatrix}$$

を仮定する(分散共分散行列は**対称行列**である事に注意)

パネルデータの分析

一般化最小二乗法を理解するには若干の行列に関する知識が必要である(とはいえ結果だけを知っておけばよいので難しくはない)。以下の事実が重要である。

「分散共分散行列 $\mathbf{\Omega}$ ($N\times N$行列) は正定値対称行列であるので、逆行列 $\mathbf{\Omega}^{-1}$が存在し、さらに $\mathbf{\Omega}^{-1}=\mathbf{\Psi}\mathbf{\Psi}^{\mathrm{T}}$となる$N\times N$行列 $\mathbf{\Psi}$ が存在する。$\mathbf{\Psi}$ は逆行列 $\mathbf{\Psi}^{-1}$を持つ。」

(演習問題2参照)

パネルデータの分析

このような行列 $\boldsymbol{\Psi}^{\mathrm{T}}$ を先ほどの回帰式の左から掛けて、

$$\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{Y} = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}$$

を得る。$\tilde{\mathbf{Y}} = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{Y}$, $\tilde{\mathbf{X}} = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{X}$, $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}$, とおくと上記の式は

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

という回帰式を表している。

パネルデータの分析

この回帰式の中の誤差項 $\tilde{\boldsymbol{\varepsilon}}$ の分散共分散行列は

$$
\begin{aligned}
E[\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}] = E[\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\Psi}] = \boldsymbol{\Psi}^{\mathrm{T}}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]\boldsymbol{\Psi}
&= \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}})^{-1}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{\Psi}^{\mathrm{T}})^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi} \\
&= \mathbf{I}_N
\end{aligned}
$$

となり、**相関なし均一分散の最小二乗法の仮定を満たす！**

パネルデータの分析

よって

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

に最小二乗法を適用すれば $\boldsymbol{\beta}$ の効率的な推定量が得られる。

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= (\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{Y}} \\ &= (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{Y} \\ &= (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Y}\end{aligned}$$

この $\hat{\boldsymbol{\beta}}_{GLS}$ を**一般化最小二乗推定量(generalized least square estimator; GLS)** という。

パネルデータの分析

■ 実行可能な一般化最小二乗法

一般化最小二乗推定量を計算するには $\mathbf{\Omega}^{-1}$ が分かっていなければならないが、実際にこれが前もってわかっている事は、いくつかの特殊ケースを除き、ほとんどない。このような場合 $\mathbf{\Omega}^{-1}$ は**その推定値** $\hat{\mathbf{\Omega}}^{-1}$ で置き換えられる。$\hat{\boldsymbol{\beta}}_{GLS}$ の$\mathbf{\Omega}^{-1}$をその推定値で置き換えたもの：

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{Y}$$

を**実行可能な一般化最小二乗法 (feasible generalized least square; FGLS)** という。$\hat{\mathbf{\Omega}}^{-1}$ の推定法についてはここでは立ち入らない。

パネルデータの分析

- パネルデータ回帰モデルのGLSによる推定

次のパネルデータ回帰式をGLSで推定する。

$$Y_{it} = \beta_1 X_{1,it} + \ldots + \beta_K X_{K,it} + u_{it}\ , \quad i = 1,\ldots, N,\ t = 1,\ldots,T$$

推定するパラメーターは $\boldsymbol{\beta} = [\,\beta_1,\ldots,\beta_K\,]^{\mathrm{T}}$ のK個である。
Y_{it} , $X_{k,it}$, u_{it}, についてまず i を固定して Y_{it}, $t = 1,\ldots,T$ まで並べる。

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} X_{1,i1} & \cdots & X_{K,i1} \\ X_{1,i2} & \cdots & X_{K,i2} \\ \vdots & \vdots & \vdots \\ X_{1,iT} & \cdots & X_{K,iT} \end{bmatrix}, \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix},$$

パネルデータの分析

すると先ほどの回帰式は

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i\ ,\ \ i = 1,\ldots, N$$

と表す事ができる。この時、$\mathbf{u}_i$ の分散共分散行列は

$$\mathrm{var}(u_{it}) = \sigma_\varepsilon{}^2 + \sigma_\alpha{}^2,\ \ \mathrm{cov}(u_{it}, u_{js}) = \begin{cases} 0 & \text{for } i \neq j \\ \sigma_\alpha^2 & \text{for } i = j \text{ and } t \neq s \end{cases}$$

より

$$E[\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}] = \boldsymbol{\Sigma}^* = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\mu^2 & \sigma_\mu^2 & \ldots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\varepsilon^2 + \sigma_\mu^2 & \ldots & \sigma_\mu^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \ldots & \sigma_\varepsilon^2 + \sigma_\mu^2 \end{bmatrix}$$

となる。

パネルデータの分析

さらに $\mathbf{Y} = [\mathbf{Y}_1{}^{\mathrm{T}}, \mathbf{Y}_2{}^{\mathrm{T}}, \ldots, \mathbf{Y}_N{}^{\mathrm{T}}]^{\mathrm{T}}$ とすると、先ほどの N 個の回帰式の行列表現はまとめて

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

と表す事ができる。ここで

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix},$$

であり、$\boldsymbol{\beta}$ は先ほどと同じである。

パネルデータの分析

この時、誤差項 **u** の分散共分散行列は

$$E[\mathbf{u}\,\mathbf{u}^{\mathrm{T}}] = \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}^* & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}^* & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma}^* \end{bmatrix}$$

となり、これは相関ゼロ、均一分散の仮定を**満たさない**。
よって、GLSで推定する。

パネルデータの分析

このモデルのGLSは

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

で与えられる。また $\boldsymbol{\Sigma}^{-1}$ は

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{*-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{*-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}^{*-1} \end{bmatrix}$$

で与えられる。$\boldsymbol{\Sigma}^{*-1}$ は$\boldsymbol{\Sigma}^{*}$の逆行列である。

パネルデータの分析

実際にはΣ^{-1} は直接観測できないので、その推定値で置き換えたFGLS

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}$$

を使用する。

演習問題(パネルデータの分析)

問題1

撹乱項 u_i を 個人効果 μ_i とそれ以外の効果 ε_{it} に分解したとき、μ_i と ε_{it} が(A1) ~ (A3) の仮定を満たす時に u_i は

$$E(u_{it}) = 0,\ \ \mathrm{var}(u_{it}) = \sigma_\alpha{}^2 + \sigma_\varepsilon{}^2,$$

$$\mathrm{cov}(u_{it}, u_{js}) = \begin{cases} 0, & i \neq j, \\ \sigma_\alpha^2, & t \neq s \end{cases}$$

を満たす事を確認しなさい。

演習問題(パネルデータの分析)

問題2

時系列分析の講義で三角分解 $\mathbf{\Omega} = \mathbf{ADA}^{\mathrm{T}}$ が出てきたが、この時 $\mathbf{\Psi}$ は $\mathbf{A}$ と $\mathbf{D}^{1/2}$ を用いてどのように表す事ができるか答えなさい。また

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

について、$\mathbf{\Psi}$ の具体的な形を求めなさい。