

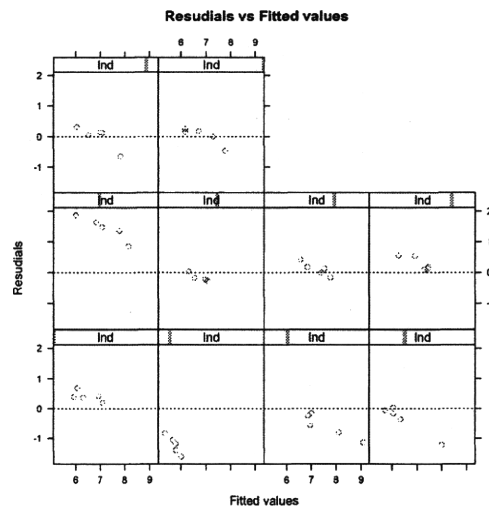
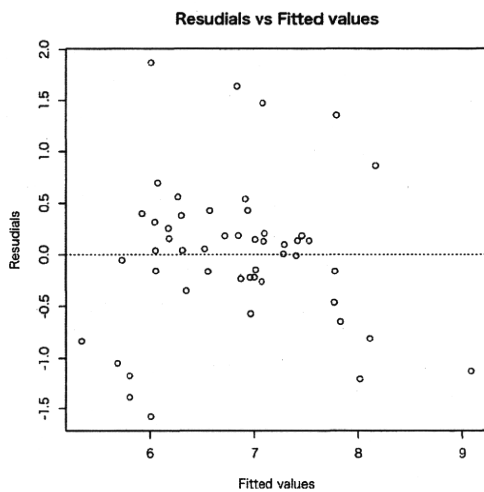
## 1.2 回帰モデルの残差

データセット A の変数 Y を変数 X で説明する回帰モデルについて、モデルにより推定された値に対する残差をプロットしてみよう。

モデルにより推定された値に対する残差をプロット

```
plot(predict(lm.ols), residuals(lm.ols),
      main = "Residuals vs Fitted values",
      xlab = "Fitted values", ylab = "Residuals")
abline(h=0, lty=3)

library(lattice)
xyplot(residuals(lm.ols) ~ predict(lm.ols) | Ind, data=Ydf,
       panel = function(x,y){
         panel.xyplot(x,y)
         panel.abline(h=0, lty=3)
       },
       main = "Residuals vs Fitted values",
       xlab = "Fitted values", ylab = "Residuals")
```



ここから、各個人別に残差を観察した場合、推定された値が大きくなるほど残差が減少する傾向が見られることがわかる。これは、観察不可能な個体の効果の存在を示唆している。

### 1.3 パネルデータの表示法

パネルデータは、同一の個体に対して複数時点での観察を行うことから、クロスセクションデータと時系列データの両方を併せ持ったデータであるといえる。したがって、パネルデータに対する回帰分析は、通常のカロスセクションデータや時系列データと異なり、変数に個体番号と時刻の2つの添字を併せ持っている。

いま、 $N$  個の個体を  $i = 1, \dots, N$  で表し、時刻を  $t = 1, \dots, T$  で表す。変数としては、被説明変数  $y_{it}$  と、これに対する  $K$  個の説明変数  $X'_{it} = [x_{1,it} \ x_{2,it} \ \dots \ x_{K,it}]$  を考える。このとき、定数項を  $\alpha$ 、定数項以外の回帰係数を  $\beta' = [\beta_1 \ \beta_2 \ \dots \ \beta_K]$ 、誤差項を  $u_{it}$  として、回帰式は、

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T$$

と書くことができる。本章では、一元配置誤差構成要素回帰モデル (One-way Error Component Regression Model) を対象とする。これは、誤差項が

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

と表されるモデルである。ここで、 $\mu_i$  は観察不可能な個体の効果、 $\nu_{it}$  は攪乱項である。

## 1.4 固定効果モデル

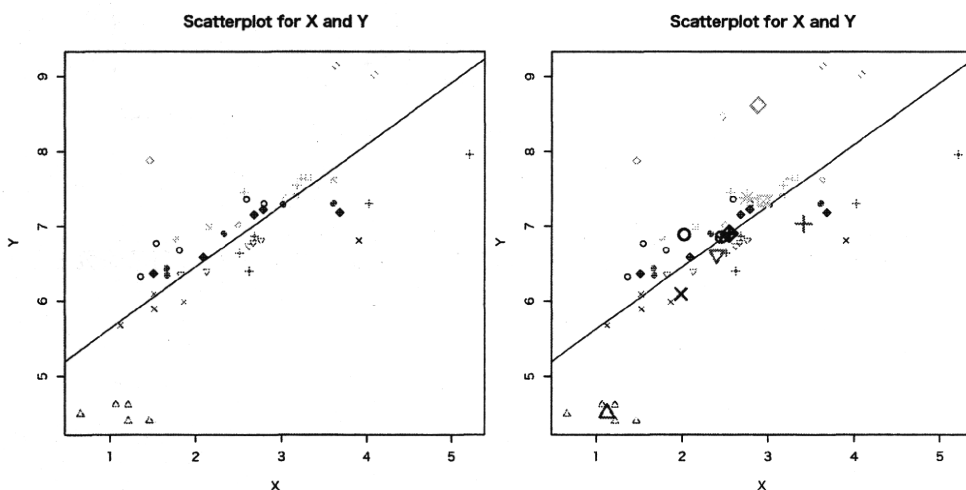
固定効果モデルとは、 $\mu_i$ : 観察不可能な個体の効果をパラメータとして推定するモデルである。すなわち、傾きは同一だが、個人毎に切片が異なる直線で回帰をするモデルといえる。推定にあたっては、まず、異なる切片の影響を除去するため、XとYから個人毎の平均値を引き去った変数を考えて、これに回帰を施すことにより傾きを推定する。

この原理を、データセットAを使ってグラフ上で考えてみよう。まず、データセットAの散布図を個人毎に色分けして描き直す。

### データセットAの散布図 (個人別)

```
plot(Ydf$X, Ydf$Y, type="p", col = Ydf$Ind, pch= Ydf$Ind,
     main = "Scatterplot for X and Y",
     xlab = "X", ylab = "Y")
abline(lm.ols)

# 個人毎の平均値を表示
Yb <- tapply(Ydf$Y, Ydf$Ind, mean)
Xb <- tapply(Ydf$X, Ydf$Ind, mean)
points(Xb, Yb, pch = seq_along(Yb), cex = 2, col = seq_along(Yb), lwd=3)
```



OLSによる傾きが個別の傾きよりやや大きいように見える。右は、個人毎に5時点のX,Yの平均値のポイントをを少し大きめのマーカーで表示したものである。個人毎の異質性を除いた傾きを推定するために、この大きなマーカーを原点に移すように移動する。

個人毎の平均値を引く

```
# 個人毎の平均値を引いたデータを作成
```

```
Yw <- Ydf$Y - Yb[Ydf$Ind]
```

```
Xw <- Ydf$X - Xb[Ydf$Ind]
```

```
plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
```

```
  main = "Scatterplot for X and Y",
```

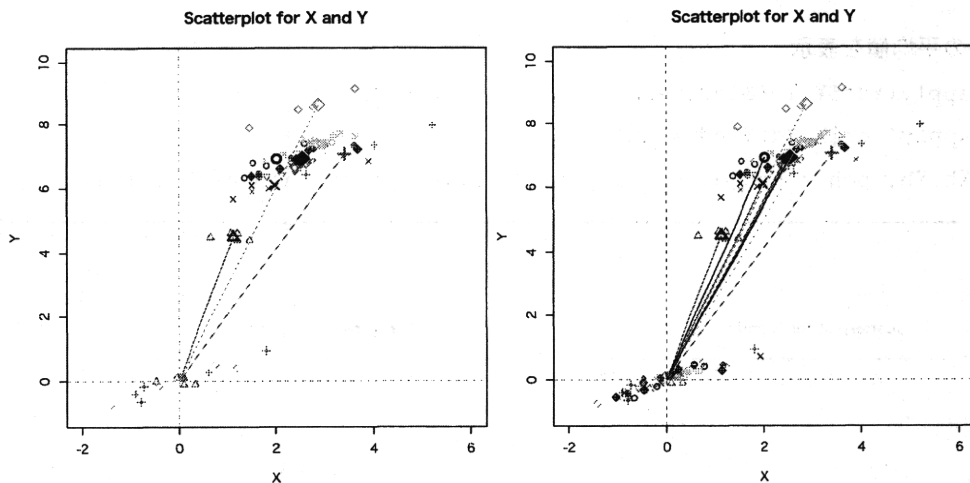
```
  xlab = "X", ylab = "Y")
```

```
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
```

```
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
```

```
abline(h=0, v=0, lty=2, lwd = 0.5)
```

```
arrows(Xb, Yb, 0, 0, col = seq_along(Yb), length = 0.1)
```



このデータに線形回帰モデルを当てはめたときの傾きが固定効果モデルの  $X$  の回帰係数となる。

## 固定効果モデル (原始的な方法)

```

lm.w <- lm(Yw ~ Xw)
summary(lm.w)

plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
     main = "Scatterplot for X and Y",
     xlab = "X", ylab = "Y")
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
abline(h=0, v=0, lty=2, lwd = 0.5)
abline(lm.w)

```

## 出力結果

```

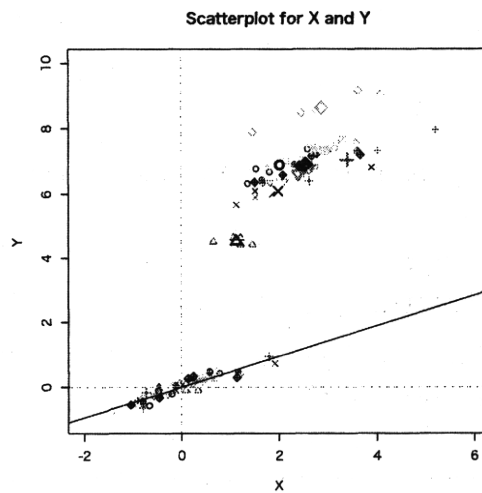
Call:
lm(formula = Yw ~ Xw)

Residuals:
    Min       1Q   Median       3Q      Max
-0.267974 -0.090150  0.001145  0.100978  0.221316

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.115e-16  1.919e-02  1.10e-14      1
Xw           4.716e-01  2.673e-02   17.64 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1357 on 48 degrees of freedom
Multiple R-squared:  0.8664, Adjusted R-squared:  0.8636
F-statistic: 311.3 on 1 and 48 DF,  p-value: < 2.2e-16

```



固定効果モデルは、図でわかるように、個人間の差異を無視して、個人内の変動に着目した推定であることから、ウィズイン (Within) 推定とも呼ばれる。これに対して、個人毎の平均値、すなわち、大きなマーカーに対して線形回帰モデルを当てはめたものを、ビトウィーン (Between) 推定量と呼ぶ。また、OLS 推定量は、全てをプールして推定を行っていることから、プーリング推定量とも呼ばれる。

#### ビトウィーン推定 (原始的な方法)

```
lm.b <- lm(Yb ~ Xb)
summary(lm.b)

plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
     main = "Scatterplot for X and Y",
     xlab = "X", ylab = "Y")
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
abline(h=0, v=0, lty=2, lwd = 0.5)
abline(lm.ols, lty=2, col=2, lwd=1.5)
abline(lm.w, lty=3, col=3, lwd=1.5)
abline(lm.b, lty=4, col=4, lwd=1.5)
legend("topleft", c("OLS(Pooled)", "Fixed Effect(Withih)", "Between"),
     lty=seq(2,4), col=seq(2,4), lwd = 1.5, cex=0.65)
```

## 出力結果

```

Call:
lm(formula = Yb ~ Xb)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0564 -0.1404 -0.0810  0.1207  1.2245

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.5860     0.8635   4.153  0.00320 **
Xb             1.3200     0.3420   3.859  0.00481 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.651 on 8 degrees of freedom
Multiple R-squared:  0.6506, Adjusted R-squared:  0.6069
F-statistic: 14.89 on 1 and 8 DF,  p-value: 0.004814

```

これらの結果から、 $X$  の回帰係数について、OLS では 0.82、固定効果モデルでは 0.47、ピトウィーン推定では 1.32 となることがわかった。実は、OLS の回帰係数は、固定効果モデルとピトウィーン推定の回帰係数の加重平均となっている。

OLS で復習した  $S_{xx}$  と  $S_{xy}$  を、パネルデータの形式で書き直し ( $S_{xx}^p$  などと表す)、ウィズイン項 ( $S_{xx}^w$  などと表す) とピトウィーン項 ( $S_{xx}^b$  などと表す) に分けると、

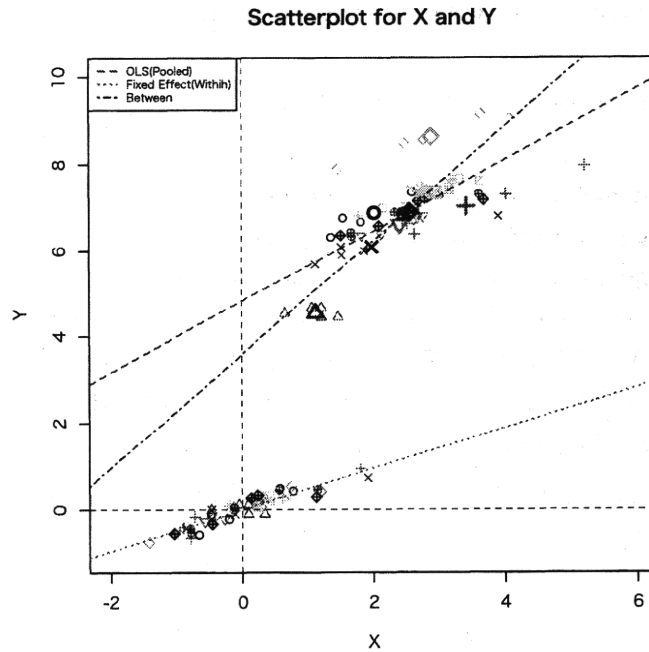
$$\begin{aligned}
 S_{xx}^p &= \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})^2 \\
 &= \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2 + \sum_{i=1}^N T(\bar{x}_i - \bar{x})^2 \\
 &= S_{xx}^w + S_{xx}^b
 \end{aligned}$$

ただし、 $\bar{x} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it}$ 、 $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$  である。 $S_{xy}^p$  についても同様の式が成り立つ。

このとき、

$$\begin{aligned}\beta^p &= \frac{S_{xy}^p}{S_{xx}^p} \\ &= \frac{S_{xy}^w + S_{xy}^b}{S_{xx}^p} \\ &= \frac{S_{xy}^w}{S_{xx}^p} \frac{S_{xx}^w}{S_{xx}^w} + \frac{S_{xy}^b}{S_{xx}^p} \frac{S_{xx}^b}{S_{xx}^b} \\ &= \frac{S_{xx}^w}{S_{xx}^p} \beta^w + \left(1 - \frac{S_{xx}^w}{S_{xx}^p}\right) \beta^b\end{aligned}$$

となるので、ウィズイン項が相対的に大きくなれば、OLS 推定は固定効果モデルに近くなる。





## OLS, 固定効果, ビトウィーン推定の関係

```
nN <- max(Ydf$Ind)
nT <- max(Ydf$time)

Sp <- cov(cbind(Ydf$X, Ydf$Y)) * (nN * nT - 1)
Sw <- cov(cbind(Xw,Yw)) * (nN * nT - 1)
Sb <- cov(cbind(Xb,Yb)) * (nN - 1) * nT

beta_p <- Sp[1,2] / Sp[1,1]
beta_w <- Sw[1,2] / Sw[1,1]
beta_b <- Sb[1,2] / Sb[1,1]

print(c(beta_p, beta_w, beta_b))
print(c(Sw[1,1]/Sp[1,1], Sb[1,1]/Sp[1,1]))
print(Sw[1,1]/Sp[1,1] * beta_w + Sb[1,1]/Sp[1,1] * beta_b)
```

## 出力結果

```
> print(c(beta_p, beta_w, beta_b))
[1] 0.8218060 0.4715958 1.3199901
> print(c(Sw[1,1]/Sp[1,1], Sb[1,1]/Sp[1,1]))
[1] 0.5872082 0.4127918
> print(Sw[1,1]/Sp[1,1] * beta_w + Sb[1,1]/Sp[1,1] * beta_b)
[1] 0.821806
```

ここまで、固定効果モデルについて、その理論的背景を確認するために原始的な方法を用いて説明を行ったが、OLS 推定に `lm` 関数があったように、パネル推定のためにも `plm` というライブラリがある。そこで、次にデータセット A の固定効果モデル等について、`plm` ライブラリの関数を用いて分析する方法を示す。なお、分析に当たって、パッケージ `plm` 及びそれに依存するパッケージを全てインストールしておくことが必要となる。

#### plm 関数による推定 (OLS)

```
library(plm)

plm.p <- plm(Y ~ X, data = Ydf, model = "pooling")
summary(plm.p)
```

#### 出力結果

```
Oneway (individual) effect Pooling Model

Call:
plm(formula = Y ~ X, data = Ydf, model = "pooling")

Balanced Panel: n=10, T=5, N=50

Residuals :
  Min. 1st Qu.  Median 3rd Qu.  Max.
-1.5900 -0.2610  0.0475  0.2990  1.8700

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  4.80749    0.28875  16.649 < 2.2e-16 ***
X              0.82181    0.11001   7.470 1.405e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    55.131
Residual Sum of Squares: 25.494
R-Squared      : 0.53758
Adj. R-Squared : 0.51608
F-statistic: 55.8014 on 1 and 48 DF, p-value: 1.4046e-09
```

## plm 関数による推定 (固定効果モデル)

```
plm.w <- plm(Y ~ X, data = Ydf, model = "within")
summary(plm.w)
summary(fixef(plm.w))
```

## 出力結果

```
> summary(plm.w)
Oneway (individual) effect Within Model

Call:
plm(formula = Y ~ X, data = Ydf, model = "within")

Balanced Panel: n=10, T=5, N=50

Residuals :
  Min.  1st Qu.  Median  3rd Qu.  Max.
-0.26800 -0.09020  0.00114  0.10100  0.22100

Coefficients :
  Estimate Std. Error t-value Pr(>|t|)
X 0.471596   0.029652  15.904 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.6146
Residual Sum of Squares: 0.88363
R-Squared      : 0.86641
  Adj. R-Squared : 0.6758
F-statistic: 252.944 on 1 and 39 DF, p-value: < 2.22e-16

> summary(fixef(plm.w))
  Estimate Std. Error t-value Pr(>|t|)
1  5.933055  0.090069  65.872 < 2.2e-16 ***
2  3.995151  0.075093  53.202 < 2.2e-16 ***
3  5.422618  0.121472  44.641 < 2.2e-16 ***
4  5.155170  0.089413  57.655 < 2.2e-16 ***
5  7.256438  0.108808  66.690 < 2.2e-16 ***
6  5.479342  0.097967  55.930 < 2.2e-16 ***
7  5.952115  0.110245  53.990 < 2.2e-16 ***
8  6.074880  0.105820  57.408 < 2.2e-16 ***
9  5.698762  0.101150  56.340 < 2.2e-16 ***
10 5.693797  0.099164  57.418 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## plm 関数による推定 (ピトウィーン推定)

```
plm.b <- plm(Y ~ X, data = Ydf, model = "between")
summary(plm.b)
```

## 出力結果

Oneway (individual) effect Between Model

Call:

```
plm(formula = Y ~ X, data = Ydf, model = "between")
```

Balanced Panel: n=10, T=5, N=50

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-1.060	-0.140	-0.081	0.121	1.220

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	3.58604	0.86351	4.1529	0.003196 **
X	1.31999	0.34204	3.8592	0.004814 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 9.7033

Residual Sum of Squares: 3.3908

R-Squared : 0.65055

Adj. R-Squared : 0.52044

F-statistic: 14.8934 on 1 and 8 DF, p-value: 0.0048144

## 1.5 ランダム効果モデル

固定効果モデルでは個人間の差異をパラメータとして推定するため、個体数が多い場合、推定しなければならないパラメータ数が大きくなることがある。しかしながら、分析上、より関心があるのは、目的変数に対する説明変数の影響、データセット A の例でいえば、変数  $x$  の回帰係数であることが往々にしてある。このような場合、個人間の差はある確率分布に従っていると考え、個別にパラメータ推定をしないという考え方もできる。

そこで、回帰式の誤差項  $u_{it} = \mu_i + \nu_{it}$  において、 $\nu_{it}$  だけでなく、 $\mu_i$  も確率変数と考え、 $\mu_i$  は平均 0、分散  $\sigma_\mu^2$  の、 $\nu_{it}$  は平均 0、分散  $\sigma_\nu^2$  の、互いに独立な同一確率分布に従うとする。

すると、誤差項  $u_{it}$  の分散・共分散行列は以下ようになる。

$$\text{cov}(u_{it}, u_{js}) = \begin{cases} \sigma_\mu^2 + \sigma_\nu^2 & (i = j \text{ and } t = s) \\ \sigma_\mu^2 & (i = j \text{ and } t \neq s) \\ 0 & (\text{otherwise}) \end{cases}$$

これをランダム効果モデルと呼ぶ。この推定には一般化最小二乗法を用いることができるが、データセット A において行った、説明変数が一つのケースについては、ランダム効果モデルで関心のある推定量である回帰直線の傾き  $\beta_r$  は、再び、固定効果モデルの傾き  $\beta^w$  とピトウィーン推定の傾き  $\beta^b$  の加重平均として、以下のように表される。

$$\beta^r = W\beta^w + (1 - W)\beta^b$$

ただし、 $W = \frac{S_{xx}^w}{S_{xx}^w + \phi^2 S_{xx}^b}$  であり、この式の中に現れる  $\phi^2$  は、 $\nu_{it}$  の分散  $\sigma_\nu^2$  と、ピトウィーン推定における誤差項の分散に時点数  $T$  を乗じた  $\sigma_i^2$  の比、すなわち、 $\phi^2 = \frac{\sigma_\nu^2}{\sigma_i^2}$  である。

なお、この式から、 $\phi^2 = 0$  のとき  $\beta^r$  は  $\beta^w$  に一致し、 $\phi^2 = 1$  のときは  $W = \frac{S_{xx}^w}{S_{xx}^w + S_{xx}^b}$  となることから、1.4 節の結果からこれは OLS 推定量  $\beta^p$  に一致することもわかる。

さて、具体的な推定を行う。ここまでで既に  $S_{xx}^w$  と  $S_{xx}^b$  は求めてあるので、あとは、 $\phi^2$  の推定値さえあれば上は計算可能である。

$\sigma_\nu^2$  の推定量  $\hat{\sigma}_\nu^2$  には、固定効果モデルで行った回帰式の誤差分散の推定量を用いることができる。これは、固定効果モデルの回帰式（原点に移動をした後のもの）の残差  $e_{it}$  の二乗平方和を自由度 ( $NT - N - K = NT - N - 1 = 39$ ) で割ることによって推定される。すなわち、 $\hat{\sigma}_\nu^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2}{NT - N - K}$  である。

また、 $\sigma_\mu^2$  は、ピトウィーン推定の誤差分散の  $T$  倍として定義される  $\sigma_i^2$  と、 $\sigma_\nu^2$  から、 $\sigma_\mu^2 = \frac{\sigma_i^2 - \sigma_\nu^2}{T}$  として求めることができる。

$\sigma_i^2$  の推定量  $\hat{\sigma}_i^2$  は、ピトウィーン推定の誤差分散の推定量の  $T$  倍であるから、その平方根を 5 で割ったものは既にピトウィーン推定の出力結果に標準偏差の推定量として示されている (Residual standard error: 0.651)。式で表せば、 $\hat{\sigma}_i^2 = T \frac{\sum_{i=1}^N e_i^2}{N - K - 1}$  である。

以上を用いてランダム効果モデルの推定を行ってみると以下ようになる。

## ランダム効果モデルの推定

```
K <- 1
sgm2_nu <- sum(residuals(lm.w)^2)/(nN * nT - nN - K)
sgm2_l <- sum(residuals(lm.b)^2)/(nN - K - 1) * nT
sgm2_mu <- (sgm2_l - sgm2_nu) / nT
print(c(sgm2_nu, sgm2_mu))

phi2 <- sgm2_nu / sgm2_l
W <- Sw[1,1] / (Sw[1,1] + phi2 * Sb[1,1])
beta_r <- W * beta_w + (1 - W) * beta_b
print(beta_r)
```

## 出力結果

```
> print(c(sgm2_nu, sgm2_mu))
[1] 0.02265726 0.41931510

> print(beta_r)
[1] 0.4779245
```

plm関数はランダム効果の推定も可能である。以下はそのコードと推定結果である。

plm関数による推定 (ランダム効果モデル)

```
plm.r <- plm(Y ~ X, data = Ydf, model = "random")
summary(plm.r)
```

出力結果

Oneway (individual) effect Random Effect Model

(Swamy-Arora's transformation)

Call:

```
plm(formula = Y ~ X, data = Ydf, model = "random")
```

Balanced Panel: n=10, T=5, N=50

Effects:

	var	std.dev	share
idiosyncratic	0.02266	0.15052	0.051
individual	0.41932	0.64755	0.949

theta: 0.8966

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-0.43500	-0.08180	-0.00466	0.11100	0.33800

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t )
(Intercept)	5.650616	0.229560	24.615	< 2.2e-16 ***
X	0.477925	0.031073	15.381	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 7.1333

Residual Sum of Squares: 1.2032

R-Squared : 0.83132

Adj. R-Squared : 0.79807

F-statistic: 236.563 on 1 and 48 DF, p-value: < 2.22e-16

## 1.6 具体的な分析例

ここでは、Balatagi (2005) において用いられている、Grunfeld(1958) の投資関数による分析例を示すとともに、推定法の選択に関する仮説検定について補足する。このモデルは以下のように記述される。

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

$I_{it}$ : 会社  $i$  の  $t$  年における投資 (real gross investment)

$F_{it}$ : 発行済株式数 (real value of the firm(shares outstanding))

$C_{it}$ : 資本金 (capital stock)

使用するのは plm パッケージに含まれる "Grunfeld" というデータであり、これに、プーリング推定、固定効果推定、ビトウィーン推定、ランダム効果推定を適用してみよう。

### Grunfeld の投資関数による分析例

```
data("Grunfeld", package = "plm")

grun.p <- plm(inv ~ value + capital, data = Grunfeld, model = "pooling")
grun.w <- plm(inv ~ value + capital, data = Grunfeld, model = "within")
grun.b <- plm(inv ~ value + capital, data = Grunfeld, model = "between")
grun.r <- plm(inv ~ value + capital, data = Grunfeld, model = "random")

result <- rbind(
  summary(grun.p)$coef[2:3,1],
  summary(grun.b)$coef[2:3,1],
  summary(grun.w)$coef[,1],
  summary(grun.r)$coef[2:3,1])
dimnames(result)[[1]] <- c("OLS", "Between", "Within", "Random")
print(result)
```

### 出力結果

	value	capital
OLS	0.1155622	0.23067849
Between	0.1346461	0.03203147
Within	0.1101238	0.31006534
Random	0.1097812	0.30811298



どの推定法を選択すべきかの検定については、以下のような考え方による。

1. 固定効果推定がプーリング推定に対して正当化されるか: F Test
2. ランダム効果推定がプーリング推定に対して正当化されるか: LM Test
3. 固定効果推定がランダム効果推定に対して正当化されるか: Hausman Test

#### 推定法選択のための検定

```
pFtest(grun.w, grun.p)
plmtest(grun.p, effect = "individual", type = "bp")
phtest(grun.w, grun.r)
```

#### 出力結果

```
> pFtest(grun.w, grun.p)

F test for individual effects

data:  inv ~ value + capital
F = 49.1766, df1 = 9, df2 = 188, p-value < 2.2e-16
alternative hypothesis: significant effects

> plmtest(grun.p, effect = "individual", type = "bp")

Lagrange Multiplier Test - (Breusch-Pagan)

data:  inv ~ value + capital
chisq = 798.1615, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects

> phtest(grun.w, grun.r)

Hausman Test

data:  inv ~ value + capital
chisq = 2.3304, df = 2, p-value = 0.3119
alternative hypothesis: one model is inconsistent
```

最初の F Test の結果から、固定効果推定はプーリング推定に対して正当化され、二番目の LM Test の結果からランダム効果推定はプーリング推定に対して正当化される。3番目の Hausman Test から固定効果推定は正当化されないため、ランダム効果推定が選択されることとなる。

## 1.7 出生児縦断調査への応用例

次に示したのは、北村 (2009) による、出生児縦断調査において、固定効果モデル・ランダム効果モデルを身長・体重の成長へ応用した例である。

表 11 体重の成長に関するパネル推定 (全体)

Dependent Variable: lnbdywht	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.002	1118.78	0.002	335.89	0.002	1238.19	0.002	1204.14
survivalday Sq	-6.25E-07	-683.11	-6.85E-07	-192.14	-6.24E-07	-751.33	-6.21E-07	-726.76
lnkosodate	-0.008	-15.09	0.001	1.20	-0.012	-22.88	-0.169	-28.81
_cons	1.176	1317.47	1.131	531.96	1.178	1311.98	1.183	1363.12
<b>Diagnostics</b>								
Number of observation	207034		207034		207034		207034	
Number of groups			46626		46626		46626	
R-sq within			0.952		0.954		0.954	
between			0.900		0.895		0.895	
overall			0.932		0.943		0.933	
F test that all $\mu_i=0$							F(46625, 160405)=2.18	
sigma_u							0.079	
sigma_e							0.154	
rho							0.210	
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 17552.43 Prob > chi2 = 0.0000			
Hausman Test							chi2(2) = 378.51 Prob > chi2 = 0.0000	

表 12 身長の成長に関するパネル推定 (全体)

Dependent Variable: lnbdyht	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.001	1322.75	0.001	420.54	0.001	1444.18	0.001	1401.47
survivalday Sq	-2.34E-07	-741.42	-2.57E-07	-222.63	-2.33E-07	-804.92	-2.32E-07	-778.47
lnkosodate	-0.003	-19.75	0.000	0.61	-0.005	-27.44	-0.007	-34.55
_cons	3.918	3.900	3.900	5709.89	3.917	12799.98	3.919	13021.36
<b>Diagnostics</b>								
Number of observation	201384		201384		201384		201384	
Number of groups			46576		46576		46576	
R-sq within			0.971		0.972		0.972	
between			0.940		0.939		0.939	
overall			0.959		0.960		0.960	
F test that all $\mu_i=0$							F(46575, 154805)=2.02	
sigma_u							0.025	
sigma_e							0.053	
rho							0.185	
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 13964.95 Prob > chi2 = 0.0000			
Hausman Test							chi2(2) = 517.03 Prob > chi2 = 0.000	

## 参考文献

- [1] Balatagi, B. H. (2005) *Econometric Analysis of Panel Data*: John Wiley and Sons, Ltd.
- [2] Croissant, Y. and G. Millo (2008) “Panel Data Econometrics in R: The plm Package”, *Journal of Statistical Software*, Vol. 27, No. 2.
- [3] 樋口美雄, 太田清, 新保一成 (2006) 『入門パネルデータによる経済分析』, 日本評論社, 東京.
- [4] 北村行伸 (2005) 『パネルデータ分析』, 岩波書店, 東京.
- [5] —— (2009) 「21世紀出生児縦断調査に基づく子供の成長パターンの測定 (III)」, 金子隆一 (編) 『厚生労働科学研究費「パネル調査(縦断調査)に関する統合的高度統計分析システムの開発研究」平成20年度総括研究報告書』, pp.127-151.

## 第2章

# ダイナミックパネル分析

### 2.1 はじめに

経済現象は基本的には経済主体がダイナミックな枠組みの中で、最適化行動を行った結果であるという認識から、最近の経済学は、異時点間の資源分配の最適化を分析の中心にして、投資、消費、雇用、金融政策、財政政策などの議論が組み立てられている。パネルデータを用いる最大のメリットの一つに、同一経済主体の異時点間の変動、すなわち動学的最適化をデータとして捉え、それを実証的に検証できるということがある。個別経済主体の初期値を知りダイナミックな変動過程（運動方程式）を知ることができれば、将来の変動や政策反応を予測できることになる。これがパネルデータを経済学者が利用したがる大きな理由になっている\*1。

パネルデータの動学的側面については、Balestra and Nerlove (1966) など1960年代より意識されてきたことではあるが、1980年代の時系列分析の発展を受けて、本格的に進展してきた。とりわけ動学的最適化にマッチした形で誕生してきた一般化積率法 (GMM) が Arellano and Bond (1991) によってパネルデータ分析に導入されて以来、急速な発展を遂げている。

本章ではダイナミック・パネルデータ分析の主要な結果をサーベイしているが、限られた紙幅では限定的なものにならざるを得ない。ダイナミック・パネルデータ分析の理論的側面について、さらに知りたい方は Arellano (2003) が包括的な参考文献となっているので参照されたい。

また、本書では全体としてクロスセクション方向の  $N$  が大きく、時系列方向の  $T$  が短いマイクロ・パネルデータを扱っており、 $N$  も  $T$  も大きいマクロ・パネルデータについては扱ってこなかった。この分野も急速に研究が進んでおり、実証上の応用も増えている。それらの研究をフォローするのは不可能に近いが、Smith and Fuerter (2004) が現在のところ最も包括的なサーベイになっている。

さらに、生存時間解析 (サバイバル分析) あるいはデュレーション・モデルとして知られている動学分析は医学、生物学を中心とした自然科学の分野で広く応用されているし、政治学、社会学の分野を中心に社会科学の分野でも最近利用されるようになってきた。これについては、日本語で読

---

\*1 本章は北村 (2005) 『パネルデータ分析』(岩波書店) の第4章に基づいている。詳細については北村 (2005) 第4章を参照されたい。実証分析は新たに『21世紀出生児縦断調査』を用いて行っている。