

上の図では、 X と Y の値によるプロットならびに X と Y の各値について個体内平均値との差分をとった値（偏差）によるプロットをそれぞれ示しており、後者に OLS 直線を当てはめている。

固定効果モデルは、図でわかるように、個体間の差異を無視して、個体内の変動に着目した推定であることから、Within(ウィズイン) 推定とも呼ばれる。これに対して、個体毎の平均値、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、Between(ビトウィーン) 推定と呼ぶ。なお、繰り返しとなるが、各個体についての全データをプールして OLS を得る場合はプーリング推定となる。

Between 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、プールド OLS では 0.82、固定効果モデルでは 0.47、Between 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 Between 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)、Within 項 (S_{xx}^w などと表す) と Between 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、

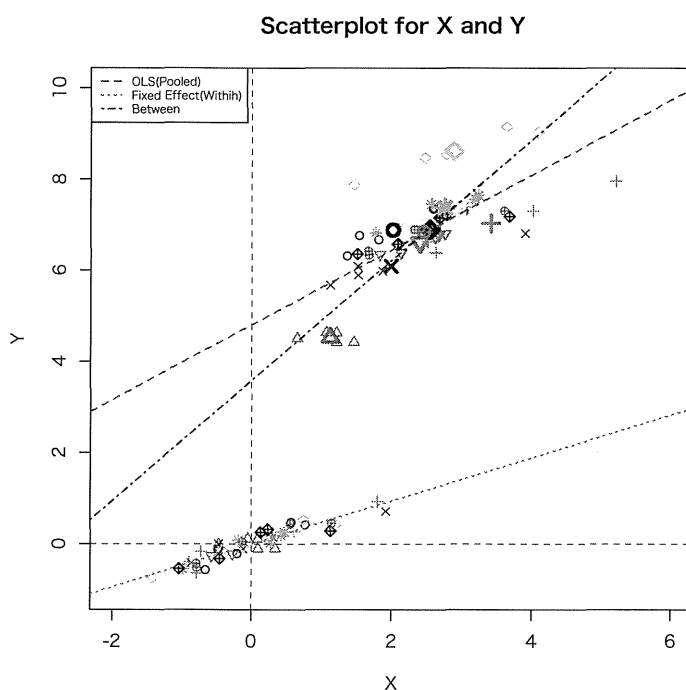
$$\beta^p = \frac{S_{xy}^p}{S_{xx}^p} \quad (5.1)$$

$$= \frac{S_{xy}^w + S_{xy}^b}{S_{xx}^p} \quad (5.2)$$

$$= \frac{S_{xy}^w}{S_{xx}^p} \frac{S_{xx}^w}{S_{xx}^w} + \frac{S_{xy}^b}{S_{xx}^p} \frac{S_{xx}^b}{S_{xx}^b} \quad (5.3)$$

$$= \frac{S_{xx}^w}{S_{xx}^p} \beta^w + \left(1 - \frac{S_{xx}^w}{S_{xx}^p}\right) \beta^b \quad (5.4)$$

となるので、Within 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。



なお、プールド OLS、固定効果モデル、そして Between 推定の回帰直線を示したものが上図である。推定方法によって、回帰直線の傾きには大きな差があることが確認できる。

OLS, 固定効果, Between 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

```

上の出力結果は、式 5.4 で示したプールド OLS、固定効果モデル、Between 推定の回帰係数の関係を確認したものである。出力結果の一つ目はそれぞれの回帰直線の傾きである β^p , β^w , β^b を示しており、二つ目は加重平均のウェイトとなる $\frac{S_w}{S_{xx}}$ と $\frac{S_b}{S_{xx}}$ が示されている。三つ目はこのウェイトを使って β^w と β^b を加重平均した結果が示されており、 β^p と一致していることが確認できる。

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 (Between 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

5.5 ランダム効果モデル

固定効果モデルでは個体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、個体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、 ν_{it} だけでなく、 μ_i も確率変数と考え、 μ_i は平均 0、分散 σ_μ^2 の、 ν_{it} は平均 0、分散 σ_ν^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・共分散行列は以下のようになる。

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

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 (GLS: generalized least square) 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 β^r は、再び、固定効果モデルの傾き β^w と Between 推定の傾き β^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 ϕ^2 は、 ν_{it} の分散 σ_ν^2 と、Between 推定における誤差項の分散に時点数 T を乗じた σ_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 β^r は β^w に一致し、 $\phi^2 = 1$ のときは $W = \frac{S_{xx}^w}{S_{xx}^b}$ となることから、5.4 節の結果からこれは OLS 推定量 β^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、 ϕ^2 の推定値さえあれば計算可能である。

σ_ν^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、 σ_μ^2 は、Between 推定の誤差分散の T 倍として定義される σ_i^2 と、 σ_ν^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。

σ_i^2 の推定量 $\hat{\sigma}_i^2$ は、Between 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 Between 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のようになる。

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
```

出力結果の一つ目は、 σ_v^2 と σ_μ^2 の推定量を示したものである。また、二つ目は、 ϕ^2 から加重平均のウェイト W を推定し、これを用いてランダム効果モデルの回帰直線の傾き β_r を推定した結果である。

`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を示したものであるが、出力結果の `Coefficients` の部分を見ると、 X の係数の推定値が 0.477925 となっており、先の結果と一致していることがわかる。

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
```

5.6 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、Between 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 |

ここで、どの推定結果を採用すべきかについて、以下の検定を通じて判断することができる。

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 検定 (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: ラグランジュ乗数検定 (Lagrange Multiplier (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: ハウスマン検定 (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 検定の結果から、固定効果推定はプーリング推定に対して正当化され、二番目のラグラ

ンジュ乗数検定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番目のハウスマン検定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。

パネルデータ分析の実際においては、プールド OLS が、固定効果モデルやランダム効果モデルに比べて、データによりよく当てはまる事例は少ない。そのため、パネルデータの分析において、この3つのモデルから最終的なモデルを選択する場合、実際には固定効果モデルかランダム効果モデルかを選択することとなる。

5.7 固定効果モデルとランダム効果モデルの選択について

前節でみたように、固定効果モデルかランダム効果モデルかの選択は、両モデルのデータへの当てはまりを通じて決定することが可能であるが、両モデルの特徴は大きく異なり、どちらのモデルを選択するかによって推定可能な変数も異なる点に注意が必要である。

稲葉 (2002) の議論を参照すると、以下のように要約できる。まず、ランダム効果モデルについては、その仮定が満たされる限り、個体内の変動と個体間の変動の両方の情報を用いて推定を行い、不偏かつ有効な推定量を得ることから、統計学的にはこれを用いることが望ましいとされる。しかし、ランダム効果モデルにおいては、モデルでは説明されない個人に固有の誤差は、すべての説明変数に対して独立であるという強い仮定をおいている。このことは、従属変数の個体間の差異を説明するあらゆる要因がモデルに含まれていることを意味しており、社会調査データを用いた分析においては該当しないケースがほとんどである。この仮定が満たされないままランダム効果モデルによる推定を行うと、Between 推定量が不偏ではなくなるため、最終的に推定される GLS 推定量にもバイアスが生じる。したがって、社会調査データを用いた分析においては、実際にはランダム効果モデルの適用が限定される場合が多い。

一方、固定効果モデルでは、個体の説明変数の変化（偏差）によって、その個体の従属変数の変化（偏差）を説明している。じつは、これは時間に対して不変な個体に固有の要因、例えば、人種や性別あるいは幼少時の家庭環境等の影響を除去した上で、説明変数の効果（回帰係数）を推定していることを意味する (Wooldridge 2003)。このとき重要な点は、この「時間に対して不変な個体に固有の要因」には、調査によって観測される変数のみならず、性格や容姿、IQ、特定の物事に対する好みや考え方等のように、必ずしも調査項目として捉えられない要因も含まれるということである。したがって、モデルでは説明されない個体に固有の誤差については、それが時間によって変化しない要因によるものであれば、この影響はすでに除去されたものとして係数の推定を行うことが可能となる。これはランダム効果モデルにおける誤差の仮定と比べて、かなり緩いものであり、データへの当てはまりとしても良好な場合が多い。

モデルでは説明されない個体に固有の誤差のことを非観察要因 (unobserved factor) ともいうが、現実の世界においては、しばしばこのような非観察要因は従属変数に重要な影響を与えており、回帰分析においてこのような要因を変数として含むことができない場合には係数の推定値にバイアスがかかることが知られている (= 欠落変数バイアス (omitted variable bias))。固定効果モデルでは、時間によって変化しないと仮定できる限り、このような非観察要因に起因するバイアス

を回避した形で回帰係数を推定することができる。性格や志向の違いなどによって人の行動パターンは千差万別であるため、そうした影響を排除した上で、諸要因のもつ直接的な効果を推定できることは、固定効果モデルの大きな強みである。

一方で、固定効果モデルでは、観察を通じて変化する要因のみしか説明変数に加えることはできない。しばしば、人種や性別などの時間によって変化しない変数と従属変数との関係が、分析者の研究対象となることがある。そのような場合に、固定効果モデルでは答えを得ることができない。また、固定効果モデルでは、観察期間を通じて全く変化が生じなかった個体については、分析から除外されることから、変化が生じることがごく希な事象の分析については、安定的な推定に必要なサンプル数を得ることができない恐れがあるため不向きである。さらに、固定効果モデルでは個体内の偏差（平均からの差）は、平均値の大小にかかわらずその重みは等しいことを仮定している。これは、例えば、テストで50点から10点上昇することと、90点から10点上昇することを同列に扱うことを意味している。同様に、回帰係数で表される各説明変数の効果は、従属変数を増加させる場合にも減少させる場合にも等しく作用することを仮定している。これらの仮定が分析にあたって妥当であるのかについては、十分に検討される必要がある。

なお、固定効果モデルでは、これらの制約を部分的に緩和する方法として、以下のような方法がある。まず、時間不変の説明変数については、時間によって変化する説明変数との間に交互作用効果を取ることで、部分的にモデルに含めることができる。この場合、例えば、ある時間不変変数の値（例えば、人種）によって説明変数（スキルアップ研修の受講の有無）が従属変数（所得）に与える効果がどの程度異なるのか、といったことを推定できるようになる。ただし、時間不変変数の主効果については、モデルでは推定することはできない。次に、個体内の平均値によって異なる偏差の重みの問題については、従属変数の個体内平均値のばらつきが大きいことが問題であるならば、従属変数の個体内平均値によって標本を分割し、できるだけ同質な集団の中で固定効果モデルを実行するという方法が考えられる。また、説明変数の個体内平均値のばらつきが大きいことが問題であるならば、その説明変数について個体内平均値と個体内偏差との交互作用項をモデルに含め、個体内平均の水準によって、説明変数の効果が異なるモデルを設定するといった方法が検討される。各説明変数の効果について、正の効果と負の効果を別々に推定する方法については有田（2013）を参照されたい。

5.8 固定効果モデルとランダム効果モデルのハイブリッドモデル

二者択一のように思われる固定効果モデルとランダム効果モデルであるが、この両者の特性を併せもつハイブリッドモデルを推定することが可能である（Allison 2009）ハイブリッドモデルにおいては、時間不変の変数と時間によって変化する変数の両方をモデルに含めることができる。しかも、時間不変の変数についてはランダム効果係数を得て、時間によって変化する変数については固定効果係数を得ることができる。このとき得られる固定効果係数は、同じサンプルと説明変数を用いて固定効果モデルを行った場合と全く同じ値と標準誤差を得る。そのため、ハイブリッドモデルは、時間不変変数を含むことが可能な固定効果モデルと解釈することも可能である。

ハイブリッドモデルの分析の手順は以下である（Allison 2009）。

- [1] 時間によって変化する変数 X について、個体内平均との差である個体内偏差をとって説明変数として用いる (→固定効果モデルと同じ)
- [2] 従属変数 Y については、偏差を取らず、そのままの値を用いる
- [3] 時間不変の変数である Z を説明変数に含める
- [4] 時間によって変化する変数 X について、個体内平均値を算出して、該当する各データレコードに貼り付け、これを説明変数に含める
- [5] OLS の代わりに、ランダム効果モデルを用いて推定を行う

この手順で分析を行うと、[1] については固定効果係数を、[3] と [4] についてはランダム効果係数を得る。このとき、[1] で得られる係数については、モデルに含まれる時間不変変数の影響を受けない (時間不変変数の specification に対して不変) という特徴がみられる。また、[4] で得られる係数については、すでに [1] の固定効果係数を得ているためそれ自体にはあまり意味はないが、この変数を含めることは、以下の2つの点で重要である。まず第1に、[3] の変数についてより良い推定値を得るために必要である。[4] の変数がない場合、モデルでは時間によって変化する説明変数 X を統制せずに、 Z の推定値を得ることになってしまう。第2に、[1] と [4] の各変数について、同一変数間でその係数を比較することでランダム効果モデルの仮定についての検証を行うことができる。もし、ランダム効果モデルの仮定が正しければ、[1] と [4] から得られる係数は同一変数については同じとなる。したがって、例えば、Wald 検定などの手法を用いれば、時間によって変化する変数 X の一つ一つについて、ランダム効果モデルの仮定が当てはまるのかを検証することができる。このようなモデルの利用についても是非検討すべきであろう。

Allison (2009) においては、ここで紹介したような線形回帰モデルのみならず、ロジスティック回帰モデル、ポワソン回帰モデル、イベントヒストリーモデル、共分散構造モデルなどにおける固定効果モデルならびにハイブリッドモデルについて解説されているため、関心のある読者はそちらを参照されたい。

5.9 出生児縦断調査への応用例

次に示したのは、Fukuda (2014) による出生児縦断調査への固定効果モデルならびにハイブリッドモデルの適用例である。

ここで示す例の従属変数は第1回から第3回調査において得られている夫の家事参加頻度（調理、食後の片付け、部屋の掃除、洗濯、ゴミ出し、日常の買い物）の6項目と育児参加頻度（食事の世話、入浴、おむつの取り替えや排泄の後始末）の3項目より算出された因子得点である。各因子得点は、第1回から第3回までの回答サンプルをプールしたデータを用いて、家事参加項目と育児参加項目のそれぞれに対して主成分分析を適用することで得ている*2。したがって、同一個人であっても調査回によって因子得点が変わる。Fukuda (2014) の分析では、これらの因子得点に対して、固定効果モデルを適用して夫の家事参加・育児参加に対する個人別の固定効果を推定しているが、ここではさらに時間によって変化しない変数を含めたハイブリッドモデルを適用した結果についても示すこととする。

説明変数は時間によって変化する変数として、夫妻の家事・育児参加についての因子得点、夫妻の調査当年の収入、妻の就業形態、夫の就業形態を使用する。時間によって変化しない変数として、夫の最終学歴、第1子の性別、結婚年、妻の第1子出生年齢を用いる。また、従属変数の時間による変化を捕捉する変数（時間変数）として調査回を用いる。

以下の表に、分析の結果を示した。分析の結果、時間によって変化する変数については、Allison (2009) にあるように固定効果モデルとハイブリッドモデルで係数ならびに標準誤差（非表示）の値が完全に一致していることから、ハイブリッドモデルにおいても成功裏に固定効果係数が推定されていることが分かる。

ハイブリッドモデルの使用によるメリットのひとつとして、時間によって変化しない変数のランダム効果を推定できることにある。この分析によって、夫の最終学歴が高く、妻の出産年齢が高い場合に夫の家事参加頻度が高いことや結婚年が最近になるほど夫の家事参加頻度が低い傾向があることなどが明らかになった。

*2 分析では第1回から第3回のすべての回について母親から回答を得ており、モデルで使用するすべての変数に欠損値がないケースに限定した、バランスパネルを用いている。

夫の家事参加・育児参加についての固定効果モデルとハイブリッドモデルの推定結果

| | 夫の家事頻度(因子得点) | | 夫の育児頻度(因子得点) | |
|----------------------------|--------------|------------|--------------|-----------|
| | ハイブリッドモデル | 固定効果モデル | ハイブリッドモデル | 固定効果モデル |
| | b | b | b | b |
| 時間変化変数 | | | | |
| 夫妻の家事・育児参加 | | | | |
| 夫の育児参加頻度(因子得点) | 0.207 *** | 0.207 *** | - | - |
| 夫の家事参加頻度(因子得点) | - | - | 0.298 *** | 0.298 *** |
| 妻の育児参加頻度(因子得点) | -0.027 *** | -0.027 *** | 0.015 ** | 0.015 ** |
| 妻の家事参加頻度(因子得点) | -0.015 *** | -0.015 *** | 0.030 *** | 0.030 *** |
| 夫妻の収入 | | | | |
| 夫の収入 | 0.000 | 0.000 | -0.030 * | -0.030 * |
| 妻の収入 | -0.017 *** | -0.017 *** | -0.008 | -0.008 |
| (妻の収入 / 夫妻の収入の合計)*10 | 0.052 *** | 0.052 *** | 0.016 | 0.016 |
| 妻の就業形態 (基準: 無職・学生) | | | | |
| 自営業・家族従業者 | -0.025 | -0.025 | -0.066 | -0.066 |
| 正規雇用 | 0.127 *** | 0.127 *** | 0.007 | 0.007 |
| 非正規雇用 | 0.061 ** | 0.061 ** | 0.000 | 0.000 |
| 夫の就業形態(基準: 正規雇用、自営業・家族従業者) | | | | |
| 無職・学生・非正規雇用 | 0.148 *** | 0.148 *** | 0.144 *** | 0.144 *** |
| 妻・妊娠 | 0.197 *** | 0.197 *** | 0.075 *** | 0.075 *** |
| 妻・出産 | -0.071 *** | -0.071 *** | 0.059 *** | 0.059 *** |
| 親と同居(基準: 非同居) | -0.380 *** | -0.380 *** | 0.013 | 0.013 |
| 保育士・保育ママ・ベビーシッターの利用 | 0.226 *** | 0.226 *** | 0.126 *** | 0.126 *** |
| 時間不変変数 | | | | |
| 調査回(基準: 第1回) | | | | |
| 第2回 | -0.127 *** | -0.127 *** | 0.210 *** | 0.210 *** |
| 第3回 | -0.184 *** | -0.184 *** | 0.265 *** | 0.265 *** |
| 夫の最終学歴(基準: 高校) | | | | |
| 中学・中卒資格の専門学校 | -0.080 *** | | -0.034 | |
| 高卒資格の専門学校・短大・高専 | 0.042 ** | | -0.026 | |
| 大学・大学院 | 0.063 *** | | -0.104 *** | |
| 第1子の性別: 女児(基準: 男児) | 0.024 * | | -0.026 ** | |
| 結婚年(基準: 1995年) | -0.017 *** | | 0.005 | |
| 妻の第1子出生年齢 (基準: 25~29歳) | | | | |
| 24歳未満 | -0.044 ** | | 0.023 | |
| 30~34歳 | 0.010 | | -0.050 *** | |
| 35歳以上 | 0.067 ** | | -0.052 * | |
| 時間変化変数の個人内平均値 | | | | |
| 夫妻の家事・育児参加 | | | | |
| 夫の育児参加頻度(因子得点) | 0.546 *** | | - | |
| 夫の家事参加頻度(因子得点) | - | | 0.500 *** | |
| 妻の育児参加頻度(因子得点) | -0.153 *** | | 0.084 *** | |
| 妻の家事参加頻度(因子得点) | 0.053 *** | | -0.081 *** | |
| 夫妻の収入 | | | | |
| ln(夫の収入) | -0.005 | | -0.009 | |
| ln(妻の収入) | -0.004 | | 0.011 | |
| (妻の収入 / 夫妻の収入の合計)*10 | -0.007 | | 0.019 | |
| 妻の就業形態 (基準: 無職・学生) | | | | |
| 自営業・家族従業者 | -0.025 | | -0.010 | |
| 正規雇用 | 0.337 *** | | -0.054 | |
| 非正規雇用 | 0.103 ** | | -0.012 | |
| 夫の就業形態(基準: 正規雇用、自営業・家族従業者) | | | | |
| 無職・学生・非正規雇用 | 0.048 | | -0.097 * | |
| 妻・妊娠 | -0.002 | | 0.135 *** | |
| 妻・出産 | -0.074 | | 0.111 | |
| 親と同居(基準: 非同居) | -0.468 *** | | 0.092 *** | |
| 保育士・保育ママ・ベビーシッターの利用 | 0.130 *** | | 0.092 ** | |
| 定数 | 0.192 * | 0.073 | -0.131 | -0.018 |
| 人-年数 | 33597 | 33597 | 33597 | 33597 |
| サンプル数 | 11199 | 11199 | 11199 | 11199 |
| 個人効果と説明変数の相関係数 | - | 0.238 | - | 0.188 |
| 個人内 R2 | 0.112 | 0.112 | 0.136 | 0.136 |
| 個人間 R2 | 0.356 | 0.293 | 0.310 | 0.288 |
| 全体 R2 | 0.308 | 0.247 | 0.258 | 0.233 |
| 個人効果によって説明されるyの分散の割合 | 0.621 | 0.700 | 0.478 | 0.576 |
| 自由度 | 38 | 11214 | 38 | 11214 |

*: p<.10, **: p<.05, ***: p<.01

さらに、ハイブリッドモデルにおいては、時間によって変化する変数について、ランダム効果モデルの仮定が当てはまるのかどうかを変数毎に検定することができる。ランダム効果モデルでは、時間によって変化する変数について、個人内平均からの偏差を用いて推定した固定効果係数と個人内平均値を用いて推定した係数とが同一であることを仮定している。そのため、Stataでは以下のコマンドを用いて Wald 検定を行い、この仮定の検定を行うことができる。

```
/* 各変数の Wald 検定 */
* 変数名_dm は個人内平均からの偏差を取った変数、変数名_m は個人内平均値を表す変数 *

foreach var in h_kaji_fac w_kaji_fac w_ikuji_fac preg ybro_now gpcore c_care w_job2
w_job3 w_job4 f_part ln_h_inc2_f ln_w_inc2_f rw_incw2_f {
test 'var'_dm 'var'_m
}

/* 時間によって変化する変数全体で仮定がみたされているか否かの Wald 検定 */

test (h_kaji_fac_dm = h_kaji_fac_m) (w_ikuji_fac_dm = w_ikuji_fac_m) (w_kaji_fac_dm =
w_kaji_fac_m) (preg_dm = preg_m) (ybro_now_dm = ybro_now_m) (gpcore_dm = gpcore_m)
(c_care_dm = c_care_m) (w_job2_dm = w_job2_m) (w_job3_dm = w_job3_m) (w_job4_dm =
w_job4_m) (f_part_dm = f_part_m) (ln_h_inc2_f_dm = ln_h_inc2_f_m) (ln_w_inc2_f_dm =
ln_w_inc2_f_m) (rw_incw2_f_dm = rw_incw2_f_m)
```

上記のコマンドを実行した結果を表したのが以下の表である。

偏差変数の係数と平均変数の係数が同じであるか否かに関する Wald 検定の結果

| | 夫の家事参加モデル | | 夫の育児参加モデル | |
|----------------------|-----------|-------|-----------|-------|
| | カイ2乗値 | P値 | カイ2乗値 | P値 |
| 夫の育児参加頻度(因子得点) | - | - | 5664.84 | 0.000 |
| 夫の家事参加頻度(因子得点) | 5664.84 | 0.000 | - | - |
| 妻の育児参加頻度(因子得点) | 262.16 | 0.000 | 83.27 | 0.000 |
| 妻の家事参加頻度(因子得点) | 36.17 | 0.000 | 96.41 | 0.000 |
| 夫の収入 | 0.08 | 0.961 | 3.95 | 0.138 |
| 妻の収入 | 8.76 | 0.013 | 2.71 | 0.258 |
| (妻の収入 / 夫妻の収入の合計)*10 | 31.29 | 0.000 | 3.89 | 0.143 |
| 妻・自営業・家族従業者 | 0.78 | 0.679 | 2.65 | 0.266 |
| 妻・正規雇用 | 94.29 | 0.000 | 2.06 | 0.357 |
| 妻・非正規雇用 | 10.95 | 0.004 | 0.08 | 0.963 |
| 夫・無職・学生・非正規雇用 | 24.43 | 0.000 | 18.66 | 0.000 |
| 妻・妊娠 | 240.05 | 0.000 | 32.52 | 0.000 |
| 妻・出産 | 26.19 | 0.000 | 14.56 | 0.001 |
| 親と同居(基準: 非同居) | 803.01 | 0.000 | 19.99 | 0.000 |
| 保育士・保育ママ・ベビーシッターの利用 | 234.28 | 0.000 | 53.65 | 0.000 |
| 全変数 | 1339.21 | 0.000 | 480.18 | 0.000 |

この検定においては、偏差変数と平均変数の係数は同一であるという帰無仮説を検定している。モデルの選択上重要であるのは、時間によって変化する変数全体で帰無仮説を棄却できるかどうかであるが (Allison 2009)、ここでは夫の家事参加・育児参加ともに帰無仮説が棄却されているた

め、ランダム効果モデルではなく固定効果モデルが選択される。この検定方法によるモデルの選択は一般的に使われるハウスマン検定の代替であるともいえる。次に、各変数の検定結果をみると、夫の家事参加モデルでは、夫の収入と妻が自営業・家族従業者であるか否かが、育児参加モデルでは、夫妻の収入に関する変数と妻の就業に関する変数において10%水準で帰無仮説が棄却できず、ランダム効果係数による解釈を許容する結果となっている。ただし、推定結果にあるように、これらの変数の係数はいずれもほぼ0であるため、係数が0であることを追認しているに過ぎない。

このようにハイブリッドモデルでは、時間によって変化しない変数の効果を固定効果係数と同時に推定することで、通常の固定効果モデルよりも多くの情報を得ることができるという点に特徴がある。稲葉(2002)が指摘するように、社会学的な分析においては、時間によって変化しない変数の水準による従属変数の相違が分析の関心となることも多いため、このようなモデルの利用も併せて検討されるべきであろう。また、時間によって変化する変数について、変数毎にランダム効果の妥当性を検討できるという点についても、従来のハウスマン検定にはない利点であると言える。

参考文献

Allison, P. D. (2009) *Fixed Effects Regression Models, Series: Quantitative Applications in the Social Sciences*, Sage Publications Inc.

有田伸(2013)「変化の向き・経路と非変化時の状態を区別したパネルデータ分析: 従事上の地位変化がもたらす所得変化を事例として」『理論と方法』vol. 28: 69-85.

Balatagi, B. H. (2005) *Econometric Analysis of Panel Data*: John Wiley and Sons, Ltd.

Croissant, Y. and G. Millo (2008) "Panel Data Econometrics in R: The plm Package", *Journal of Statistical Software*, Vol. 27, No. 2.

Fukuda, S. (2014) "Gender Equality and Transition to Second Birth in Japan", 金子隆一(編)『厚生労働科学研究費補助金「縦断および横断調査によるライフコース事象の経時変化分析と施策への対応に関する研究(平成24-政策-一般-004)」平成25年度総括研究報告書』

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.

稲葉昭英(2002)「研究動向 Pooled Time Series モデル」, 『家族社会学研究』, 第14巻, 第1号, pp. 5-10.

北村行伸(2005)『パネルデータ分析』, 岩波書店, 東京.

Wooldridge, J. M. (2003) *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson.

第 6 章

ダイナミックパネル

6.1 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。