Панельные регрессии

ЦМФ

Основные обозначения

$i \in \{1; \dots; n\}$ — объекты, $t \in \{0; \dots; T\}$ — периоды времени

$\vec{y}_i = \left(y_{i,1}, \dots, y_{i,T}\right)'$ — вектор зависимой переменной для i-го объекта

$X_i = \left(\vec{x}_{i,1}, \dots, \vec{x}_{i,T}\right)'$ — матрица значений регрессоров для i-го объекта

$\vec{x}_{i,t} = \left(x_{i,t}^{(1)}, \dots, x_{i,t}^{(k)}\right)'$ — вектор значений регрессоров для i-го объекта и периода времени t

$\vec{y}_{[nT\times 1]} = (\vec{y}_1', \dots, \vec{y}_n')'$ — объединённый вектор значений зависимой переменной

$X_{[nT\times k]} = (X_1, \dots, X_n)'$ — объединённая матрица регрессоров

Объединённая модель регрессии (pooled model)

Применяется обычная линейная регрессия без учёта панельной структуры данных:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

Предпосылки:

- $cov(\varepsilon_{i,t}, \varepsilon_{j,s}) = cov(\varepsilon_{i,t}, x_{j,s,q}) = 0$

МНК-оценки $\hat{\vec{\beta}}$ являются состоятельными и эффективными

Модель с фиксированными эффектами (fixed effect, within model)

$$y_{i,t} = \alpha_i + \vec{x}'_{i,t}\vec{\beta} + \varepsilon_{i,t} \quad (1)$$

α_i — индивидуальный эффект i-го объекта

Предпосылки:

- $cov(\varepsilon_{i,t}, \varepsilon_{j,s}) = 0$
- $E(\varepsilon_{i,t}) = 0,\ var(\varepsilon_{i,t}) = \sigma_\varepsilon^2$
- $cov(\varepsilon_{i,t}, x_{j,s,q}) = 0$

Модель с фиксированными эффектами в матричной форме

Пусть $\begin{cases} d_{i,t} = 1, & i = j \\ d_{i,t} = 0, & i \neq j \end{cases}$, тогда (1) можно переписать в виде

$$y_{i,t} = \sum_{j=1}^{n} \alpha_j d_{i,j} + \vec{x}'_{i,t}\vec{\beta} + \varepsilon_{i,t} \quad (2)$$

Пусть $D = \begin{pmatrix} \vec{\imath}_T & 0 & \dots & 0 \\ 0 & \vec{\imath}_T & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \vec{\imath}_T \end{pmatrix}$, где $\vec{\imath}_T = (1, \dots, 1)'$, и

$\vec{\alpha} = (\alpha_1, \dots, \alpha_n)'$, тогда (2) представляется в матричной форме:

$$\vec{y} = D\vec{\alpha} + X\vec{\beta} + \vec{\varepsilon} \quad (3)$$

МНК-оценки несмещённые и эффективные

Состоятельность имеет место при увеличении числа периодов T, но не достигается при росте числа объектов n

Внутригрупповое преобразование (within transformation)

Рассмотрим уравнение (1) в средних:

$$\bar{y}_i = \alpha_i + \bar{x}_i'\vec{\beta} + \bar{\varepsilon}_i, \quad (4) \quad \text{где}$$

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}\,, \;\; \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} \vec{x}_{i,t}\,, \;\; \bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}$$

Вычтем (4) из (1):

$$y_{i,t} - \bar{y}_i = \left(\vec{x}_{i,t} - \bar{x}_i\right)'\vec{\beta} + \varepsilon_{i,t} - \bar{\varepsilon}_i \quad (5)$$

В матричной форме:

$$M_D\vec{y} = M_D X\vec{\beta} + M_D\vec{\varepsilon}, \quad (6) \quad \text{где}$$

$$M_D = \boldsymbol{I}_{nT} - D(D'D)^{-1}D'$$

$$\hat{\vec{\beta}}_{FE} = (X'M_DX)^{-1}X'M_D\vec{y} \quad (7)$$

Внутригрупповые оценки и оценки индивидуальных эффектов

$$\hat{\vec{\beta}}_{FE} = (X'M_D X)^{-1} X' M_D \vec{y} \quad (7)$$

$\hat{\vec{\beta}}_{FE}$ — внутригрупповая оценка (within estimator) или оценка с фиксированным эффектов (fixed effect estimator)

Оценки индивидуальных эффектов:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i' \hat{\vec{\beta}}_{FE},\ i \in \{1; \dots n\} \quad (8)$$

Ковариационная матрица внутригрупповых оценок:

$$cov\left(\hat{\vec{\beta}}_{FE}\right) = \sigma_\varepsilon^2 (X'M_D X)^{-1},$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{nT-n-k} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t} - \bar{y}_i - \left(\vec{x}_{i,t} - \bar{x}_i\right)' \hat{\vec{\beta}}_{FE} \right)^2$$

Модель со случайными эффектами (random effect)

$$y_{i,t} = \mu + \vec{x}'_{i,t}\vec{\beta} + u_i + \varepsilon_{i,t} \quad (9)$$

Предпосылки:

- $cov(\varepsilon_{i,t}, \varepsilon_{j,s}) = 0$
- $E(\varepsilon_{i,t}) = 0,\ var(\varepsilon_{i,t}) = \sigma_\varepsilon^2$
- $cov(\varepsilon_{i,t}, x_{j,s,q}) = 0$
- $cov(u_i, u_j) = 0$
- $E(u_i) = 0,\ var(u_i) = \sigma_u^2$
- $cov(u_i, \varepsilon_{j,s}) = cov(u_i, x_{j,s,q}) = 0$

Модель со случайными эффектами в матричной форме

Пусть $\vec{\omega} = \vec{u} + \vec{\varepsilon}$, тогда модель (9) запишется в матричной форме:

$$\vec{y} = \mu\vec{\imath}_{nT} + X\vec{\beta} + \vec{\omega} \quad (10)$$

Оценки параметров:

$$\begin{bmatrix} \hat{\vec{\mu}}_{RE} \\ \hat{\vec{\beta}}_{RE} \end{bmatrix} = \left(\begin{bmatrix} \vec{\imath}'_{nT} \\ X' \end{bmatrix} (\boldsymbol{I}_n \otimes \Sigma^{-1}) [\vec{\imath}_{nT} \quad X] \right)^{-1} (\boldsymbol{I}_n \otimes \Sigma^{-1}) \vec{y}, \quad (11)$$

$\otimes$ — произведение Кронекера,

$$\Sigma^{-1} = \frac{1}{\sigma_\varepsilon^2} \left(\left(\boldsymbol{I}_T - \frac{1}{T} \vec{\imath}_T \vec{\imath}'_T \right) + \frac{1}{T} \cdot \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} \cdot \vec{\imath}_t \vec{\imath}'_t \right)$$

Межгрупповое преобразование (between transformation)

Модель в средних:

$$\bar{y}_i = \mu + \bar{x}_i'\vec{\beta} + u_i + \bar{\varepsilon}_i \quad (12)$$

$$\hat{\hat{\beta}}_B = \left(\sum_{i=1}^{n}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1} \sum_{i=1}^{n}(\bar{x}_i - \bar{x})(\bar{y}_i - y), \quad (13)$$

$$\hat{\mu}_B = \bar{y}_i - \bar{x}_i'\vec{\beta}, \quad (14) \quad \text{где}$$

$$\bar{y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{i,t}\,, \quad \bar{x} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \vec{x}_{i,t}$$

Ковариационная матрица оценки $\hat{\hat{\beta}}_{RE}$:

$$cov\left(\hat{\hat{\beta}}_{RE}\right) = \sigma_\varepsilon^2 \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\vec{x}_{i,t} - \bar{x}\right)\left(\vec{x}_{i,t} - \bar{x}\right)' + \theta^2 T \sum_{i=1}^{n}(\bar{x}_i -\right.$$

Измерители качества подгонки модели

$$R^2_{within} = cor\left(y_{i,t} - \bar{y}_i, \hat{y}_{i,t} - \hat{\bar{y}}_i\right),$$

$$R^2_{between} = cor(\bar{y}_i, \hat{\bar{y}}_i),$$

$$R^2_{overall} = cor\left(y_{i,t}, \hat{y}_{i,t}\right)$$

С помощью R^2 нельзя сравнивать разные виды моделей (например, фиксированные эффекты против случайных), но можно сравнивать модели одного вида с разными наборами регрессоров

Выбор вида модели (содержательный аспект)

Простая регрессия предполагает, что между объектами нет индивидуальных различий

Модель с фиксированными эффектами (МФЭ) — индивидуальные различия значимы, и в выборке представлена вся совокупность объектов

Случайные эффекты (МСЭ) — различия значимы, но выборка неполна

В МФЭ индивидуальный эффект коррелирован с регрессорами, в МСЭ — нет

МФЭ состоятельна в любом случае, но при отсутствии корреляции неэффективна; МСЭ в коррелированном случае несостоятельна

Выбор вида модели (проверка статистических гипотез)

Объединённая модель против фиксированных эффектов

$H_0\colon\ \alpha_1 = \cdots = \alpha_n$

Гипотеза проверяется F-тестом для модели (3)

Объединённая модель против случайных эффектов

$H_0\colon\ \sigma_u^2 = 0$

Проверяется тестом множителей Лагранжа:

$$LM = \frac{nT}{2(T-1)}\left(\frac{\sum_{i=1}^{n}\left(\sum_{t=1}^{T} e_{i,t}\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T} e_{i,t}^2} - 1\right)^2 = \frac{nT}{2(T-1)}\left(\frac{\vec{e}\,'DD'\vec{e}}{\vec{e}\,'\vec{e}} - 1\right)^2 \sim \chi^2(1)$$

Случайные эффекты против фиксированных

$H_0\colon\ cov\left(\alpha_i, \vec{x}_{j,t}\right) = 0$

Проверяется тестом Хаусмана:

$$\xi_H = \left(\hat{\hat{\beta}}_{FE} - \hat{\hat{\beta}}_{RE}\right)'\left(Cov\left(\hat{\hat{\beta}}_{FE}\right) - Cov\left(\hat{\hat{\beta}}_{RE}\right)\right)^{-1}\left(\hat{\hat{\beta}}_{FE} - \hat{\hat{\beta}}_{RE}\right) \sim \chi^2(k)$$

Панельная регрессия в R

исходные данные

```
library(plm)
data(Gasoline)

head(Gasoline)
  country year lgaspcar  lincomep      lrpmg  lcarpcap
1 AUSTRIA 1960 4.173244 -6.474277 -0.3345476 -9.766840
2 AUSTRIA 1961 4.100989 -6.426006 -0.3513276 -9.608622
3 AUSTRIA 1962 4.073177 -6.407308 -0.3795177 -9.457257
4 AUSTRIA 1963 4.059509 -6.370679 -0.4142514 -9.343155
5 AUSTRIA 1964 4.037689 -6.322247 -0.4453354 -9.237739
6 AUSTRIA 1965 4.033983 -6.294668 -0.4970607 -9.123903
```

Объединённая регрессия

```
model.formula <- lgaspcar ~ lincomep + lrpmg + lcarpcap

gas.ls <- plm(model.formula,data=Gasoline,index=c("country","year"),
              model="pooling",effect="individual")
summary(gas.ls)

Balanced Panel: n=18, T=19, N=342

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-0.3840 -0.1530 -0.0498  0.1650  0.5970

Coefficients :
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  2.391326   0.116934  20.450 < 2.2e-16 ***
lincomep     0.889962   0.035806  24.855 < 2.2e-16 ***
lrpmg       -0.891798   0.030315 -29.418 < 2.2e-16 ***
lcarpcap    -0.763373   0.018608 -41.023 < 2.2e-16 ***
---

Total Sum of Squares: 102.74   Residual Sum of Squares: 14.904
R-Squared: 0.85494             Adj. R-Squared: 0.84494
```

Модель с фиксированными эффектами

```
gas.fe <- plm(model.formula,data=Gasoline,index=c("country","year"),
              model="within",effect="individual")
summary(gas.fe)

Balanced Panel: n=18, T=19, N=342

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-0.37900 -0.03980  0.00465  0.04540  0.36300

Coefficients :
          Estimate Std. Error  t-value  Pr(>|t|)
lincomep  0.662250   0.073386   9.0242 < 2.2e-16 ***
lrpmg    -0.321702   0.044099  -7.2950 2.355e-12 ***
lcarpcap -0.640483   0.029679 -21.5804 < 2.2e-16 ***
---

Total Sum of Squares:    17.061    Residual Sum of Squares: 2.7365
R-Squared      :  0.8396           Adj. R-Squared :  0.78805
```

Извлечение индивидуальных эффектов

```
summary(fixef(gas.fe))

          Estimate Std. Error t-value  Pr(>|t|)
AUSTRIA    2.28586    0.22832 10.0115 < 2.2e-16 ***
BELGIUM    2.16555    0.21290 10.1718 < 2.2e-16 ***
CANADA     3.04184    0.21864 13.9129 < 2.2e-16 ***
DENMARK    2.38946    0.20809 11.4830 < 2.2e-16 ***
FRANCE     2.20477    0.21647 10.1851 < 2.2e-16 ***
GERMANY    2.14987    0.21788  9.8670 < 2.2e-16 ***
GREECE     2.33711    0.21488 10.8761 < 2.2e-16 ***
IRELAND    2.59233    0.24369 10.6380 < 2.2e-16 ***
ITALY      2.23255    0.23954  9.3201 < 2.2e-16 ***
JAPAN      2.37593    0.21184 11.2159 < 2.2e-16 ***
NETHERLA   2.23479    0.21417 10.4345 < 2.2e-16 ***
NORWAY     2.21670    0.20304 10.9174 < 2.2e-16 ***
SPAIN      1.68178    0.16246 10.3517 < 2.2e-16 ***
SWEDEN     3.02634    0.39451  7.6711  1.71e-14 ***
SWITZERL   2.40250    0.22909 10.4870 < 2.2e-16 ***
TURKEY     2.50999    0.23566 10.6510 < 2.2e-16 ***
U.K.       2.34545    0.22728 10.3195 < 2.2e-16 ***
U.S.A.     3.05525    0.21960 13.9128 < 2.2e-16 ***
```

Модель со случайными эффектами

```
gas.fe <- plm(model.formula,data=Gasoline,index=c("country","year"),
              model="random",effect="individual")
summary(gas.fe)

Balanced Panel: n=18, T=19, N=342

Effects:
                   var  std.dev share
idiosyncratic 0.008525 0.092330 0.182
individual    0.038238 0.195545 0.818
theta:  0.8923

Coefficients :
             Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)  1.996698   0.184326  10.8324 < 2.2e-16 ***
lincomep     0.554986   0.059128   9.3861 < 2.2e-16 ***
lrpmg       -0.420389   0.039978 -10.5155 < 2.2e-16 ***
lcarpcap    -0.606840   0.025515 -23.7836 < 2.2e-16 ***
---
R-Squared      :  0.82931       Adj. R-Squared :  0.81961
```

Проверка статистических гипотез

\# объединённая регрессия против фиксированных эффектов

```
pooltest(model.formula,data=Gasoline,index=c("country","year"),
         model="within",effect="individual")
F = 27.3352, df1 = 51, df2 = 270, p-value < 2.2e-16
alternative hypothesis: unstability
```

\# объединённая регрессия против случайных эффектов

```
plmtest(model.formula,data=Gasoline,index=c("country","year"),
        effect="individual",type="bp")
chisq = 1465.552, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

\# случайные эффекты против фиксированных

```
phtest(model.formula,data=Gasoline,index=c("country","year"),
       model=c("within","random"),effect="individual")
chisq = 302.8037, df = 3, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Динамическая модель

$$y_{i,t} = \alpha_i + \vec{x}'_{i,t}\vec{\beta} + \gamma y_{i,t-1} + \varepsilon_{i,t} \quad (15)$$

$y_{i,t-1}$ коррелированы независимо от природы эффекта

Рассмотрим модель без экзогенных переменных:

$$y_{i,t} = \alpha_i + \gamma y_{i,t-1} + \varepsilon_{i,t} \quad (16)$$

В первых разностях:

$$y_{i,t} - y_{i,t-1} = \gamma\left(y_{i,t-1} - y_{i,t-2}\right) + \varepsilon_{i,t} - \varepsilon_{i,t-1} \quad (17)$$

В уравнении (17) регрессоры и ошибки коррелированы, т.к. коррелированы $y_{i,t-1}$ и $\varepsilon_{i,t-1}$

Состоятельные оценки $\hat{\gamma}$ можно получить с помощью инструментальной переменной $y_{i,t-2}$, которая коррелирована с регрессором и не коррелирована с ошибкой

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T} y_{i,t-2}\left(y_{i,t}-y_{i,t-1}\right)}{\sum_{i=1}^{n}\sum_{t=2}^{T} y_{i,t-2}\left(y_{i,t-1}-y_{i,t-2}\right)} \quad (18)$$

Моментные тождества

Более общий случай — оценка обобщённым методом моментов на основе моментных тождеств

При $t = 2$ имеем: $E\left((\varepsilon_{i,t} - \varepsilon_{i,t-1})y_{i,t-2}\right) = E\left((\varepsilon_{i,2} - \varepsilon_{i,1})y_{i,0}\right) = 0$

При $t = 3$ имеем два тождества: $E\left((\varepsilon_{i,3} - \varepsilon_{i,2})y_{i,0}\right) = 0$

$$E\left((\varepsilon_{i,3} - \varepsilon_{i,2})y_{i,1}\right) = 0$$

При $t = T$ имеем $T - 1$ тождеств: $E\left((\varepsilon_{i,T} - \varepsilon_{i,T-1})y_{i,0}\right) = 0$

$$\dots$$

$$E\left((\varepsilon_{i,T} - \varepsilon_{i,T-1})y_{i,T-1}\right) = 0$$

Всего имеем $1 + 2 + \cdots + T - 1 = \frac{T}{2}(T + 1) - 1$ тождеств

Оценка обобщённым методом моментов

Пусть $\overrightarrow{\Delta\varepsilon_i} = \left(\varepsilon_{i,2} - \varepsilon_{i,1}, \dots, \varepsilon_{i,T} - \varepsilon_{i,T-1}\right)', Z_i =$

$$\begin{pmatrix} y_{i,0} & 0 & \cdots & 0 \\ 0 & [y_{i,0} \quad y_{i,1}] & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & [y_{i,0}, \dots, y_{i,T-2}] \end{pmatrix},$$

тогда моментные тождества можно записать в виде

$$E\left(Z_i'\overrightarrow{\Delta\varepsilon_i}\right) = E\left(Z_i'\left(\overrightarrow{\Delta y_i} - \gamma\overrightarrow{\Delta y_i}(-1)\right)\right) = 0 \quad (19)$$

Оценка:

$$\hat{\gamma}_{GMM} = \left(\left(\sum_{i=1}^{n}\overrightarrow{\Delta y_i'}(-1)Z_i\right)S\left(\sum_{i=1}^{n}Z_i'\overrightarrow{\Delta y_i}(-1)\right)\right)^{-1} \times$$
$$\left(\left(\sum_{i=1}^{n}\overrightarrow{\Delta y_i'}(-1)Z_i\right)S\left(\sum_{i=1}^{n}Z_i'\overrightarrow{\Delta y_i}\right)\right), \quad (20)$$

$$S = \left(E\left(Z_i'\overrightarrow{\Delta\varepsilon_i'}\overrightarrow{\Delta\varepsilon_i}Z_i\right)\right)^{-1}, \ \widehat{\overrightarrow{\Delta\varepsilon_i}} = \overrightarrow{\Delta e_i}$$

Модель с экзогенными регрессорами

В исходной модели

$$y_{i,t} = \alpha_i + \vec{x}'_{i,t}\vec{\beta} + \gamma y_{i,t-1} + \varepsilon_{i,t} \quad (15)$$

моментные тождества дополняются выражениями

$E\left(\vec{x}_{i,s}\varepsilon_{i,t}\right) = 0$

Переменные $\left(\vec{x}_{i,1}, \dots, \vec{x}_{i,T}\right)$ можно использовать в качестве дополнительных инструментов

Динамическая модель в R

исходные данные

```
data(Cigar)

head(Cigar)
  state year price  pop  pop16  cpi      ndi sales pimin
1     1   63  28.6 3383 2236.5 30.6 1558.305  93.9  26.1
2     1   64  29.8 3431 2276.7 31.0 1684.073  95.4  27.5
3     1   65  29.8 3486 2327.5 31.5 1809.842  98.5  28.9
4     1   66  31.5 3524 2369.7 32.4 1915.160  96.4  29.5
5     1   67  31.6 3533 2393.7 33.4 2023.546  95.5  29.6
6     1   68  35.6 3522 2405.2 34.8 2202.486  88.4  32.0
```

Пример авторегрессионной модели

```
ar.model <- dynformula(log(sales)~log(price)+log(pop)+
                       log(pop16)+pimin, lag.form=list(1,0,0,0,0))

cigar.argmm <- pgmm(ar.model,data=Cigar,index=c("state","year"),
                    gmm.inst=~log(sales)+log(price)+log(pop)+
                    log(pop16)+ndi+log(pimin),
                    lag.gmm=c(2,10))
summary(cigar.argmm)

Coefficients
                      Estimate  Std. Error z-value  Pr(>|z|)
lag(log(sales), 1)  0.69701806  0.05576996 12.4981 < 2.2e-16 ***
log(price)         -0.29907914  0.08060227 -3.7106 0.0002068 ***
log(pop)           -0.61002852  0.16616390 -3.6712 0.0002414 ***
log(pop16)          0.42526936  0.16582209  2.5646 0.0103291 *
pimin              -0.00155945  0.00071055 -2.1947 0.0281838 *

Sargan Test: chisq(1292) = 46 (p.value=1)
Autocorrelation test (1): normal = -4.945137 (p.value=3.8045e-07)
Autocorrelation test (2): normal = 2.219766 (p.value=0.013217)
Wald test for coefficients: chisq(5) = 234.1875 (p.value=< 2.22e-16)
Wald test for time dummies: chisq(28) = 844.3341 (p.value=< 2.22e-16)
```

Домашнее задание

Построить статическую и динамическую модели занятости в Великобритании

```
data(EmplUK)
```

Обосновать выбор вида модели, дать интерпретацию её коэффициентов