Uogólniona metoda najmniejszych kwadratów

Rozważmy model liniowy $(\mathbf{Y}, \mathbf{X\beta}, \sigma^2\mathbf{\Sigma})$ tzn. $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}, \operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{\Sigma}$ gdzie $\mathbf{\Sigma}$ jest znaną symetryczną dodatnio określoną macierzą.

Pokażemy, że

- Najlepszym liniowym nieobciążonym estymatorem wektora $\boldsymbol{\beta}$ jest $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$
- Macierz kowariancyjna wektora $\hat{\boldsymbol{\beta}}$ jest postaci $\operatorname{cov}\hat{\boldsymbol{\beta}} = \sigma^2(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$
- Nieobciążonym estymatorem σ^2 jest
$$s^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-k-1} = \frac{\mathbf{Y}^T(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1})\mathbf{Y}}{n-k-1}$$

Ponieważ $\mathbf{\Sigma}$ jest macierzą dodatnio określoną istnieje nieosobliwa macierz $\mathbf{P}$, taka że $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}^T$.

Mnożąc $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$ przez $\mathbf{P}^{-1}$ otrzymujemy model $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon}$, przy czym

$$E(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{P}^{-1}E(\boldsymbol{\varepsilon}) = \mathbf{0},$$

$$\operatorname{cov}(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{P}^{-1}\operatorname{cov}(\boldsymbol{\varepsilon})(\mathbf{P}^{-1})^T = \mathbf{P}^{-1}\sigma^2\mathbf{\Sigma}(\mathbf{P}^{-1})^T = \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^T(\mathbf{P}^{-1})^T =$$
$$= \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^T(\mathbf{P}^{-1})^T = \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^T(\mathbf{P}^T)^{-1} = \sigma^2\mathbf{I}.$$

Widać, że założenia twierdzenia Gaussa-Markowa są spełnione dla modelu $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon}$.

Stąd

$$\hat{\boldsymbol{\beta}} = \left((\mathbf{P}^{-1}\mathbf{X})^T(\mathbf{P}^{-1}\mathbf{X})\right)^{-1}(\mathbf{P}^{-1}\mathbf{X})^T\mathbf{P}^{-1}\mathbf{Y} = (\mathbf{X}^T(\mathbf{P}^{-1})^T\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{P}^{-1})^T\mathbf{P}^{-1}\mathbf{Y} =$$
$$= \mathbf{X}^T(\mathbf{P}^{-1})^T\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{P}\mathbf{P}^T)^{-1}\mathbf{Y} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{Y}.$$

$$\operatorname{cov}\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\operatorname{cov}(\mathbf{Y})\left((\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\right)^T =$$
$$= (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\sigma^2\mathbf{\Sigma}(\mathbf{\Sigma}^{-1})^T\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} =$$
$$= \sigma^2(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{\Sigma}^{-1})^T\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$$

$$\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X\beta} + \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}(\mathbf{X\beta} + \boldsymbol{\varepsilon}) =$$
$$\mathbf{X\beta} + \boldsymbol{\varepsilon} - \mathbf{X\beta} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1})\boldsymbol{\varepsilon} = \mathbf{M}\boldsymbol{\varepsilon},$$

gdzie $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}$

Bezpośrednim rachunkiem sprawdzamy, że $\mathbf{M}^T\mathbf{\Sigma}^{-1}\mathbf{M} = \mathbf{\Sigma}^{-1}\mathbf{M}$ i $\operatorname{tr}\mathbf{M} = n - k - 1$

Rzeczywiście

$$\operatorname{tr}\mathbf{M} = \operatorname{tr}\mathbf{I} - \operatorname{tr}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1} = n - \operatorname{tr}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X} = n - \operatorname{tr}\mathbf{I}_{k+1} = n - (k+1)$$

$$E(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\varepsilon}) = E\left(\operatorname{tr}(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}^{T}\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\varepsilon})\right) =$$
$$= E\left(\operatorname{tr}(\mathbf{M}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\right) = \operatorname{tr}(\mathbf{M}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \operatorname{tr}(\mathbf{M}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M}\sigma^2\boldsymbol{\Sigma}) = \{\mathbf{M}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M} = \boldsymbol{\Sigma}^{-1}\mathbf{M}\}$$
$$= \sigma^2\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\Sigma}) = \sigma^2\operatorname{tr}(\mathbf{M}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}) = \sigma^2\operatorname{tr}(\mathbf{M})$$

Wybrane struktury kowariancyjne

Symetryczna i dodatnio określona macierz kowariancyjna

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{m(m-1)} & \sigma_m^2 \end{bmatrix}$$

jest parametryzowana przez wektor parametrów $\theta = (\theta_1,...,\theta_k)^T$. Liczba i znaczenie parametrów zależy od rozważanej struktury kowariancyjnej. W wielu strukturach kowariancyjnych używamy dekompozycji

$\Sigma = DPD$, gdzie

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_m \end{bmatrix}$$ jest diagonalną macierzą odchyleń standardowych

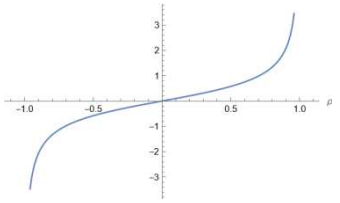
i $$P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{m(m-1)} & 1 \end{bmatrix}$$ jest macierzą korelacyjną z elementami.

z elementami $\rho_{ij} \in (-1,1)$. Jeśli struktura kowariancyjna dopuszcza różne wariancje σ_i^2 dla różnych i=1,…,m to nazywamy ją strukturą *heterogeniczną* . Zakładając $\sigma^2 = \sigma_i^2, i = 1...,m$ otrzymamy strukturę *homogeniczną*.

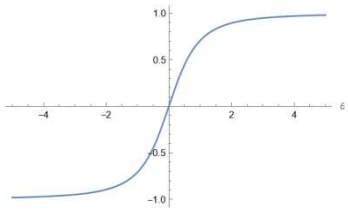
Transformacja parametrów kowariancyjnych.

Do odchyleń standardowych stosujemy transformację logarytmiczną $\log(\sigma_i), i = 1...,m$ a do parametrów korelacyjnych transformację

$$\theta = f(\rho) = sign(\rho)\sqrt{\frac{\rho^2}{1-\rho^2}}$$



$$\rho = f^{-1}(\theta) = \frac{\theta}{\sqrt{1+\theta^2}}$$



przekształcającą bijektywnie przedział (0,1) na *R*. Te transformacje są bardzo ważne gdyż do estymacji parametrów będzie można użyć procedury estymacji bez ograniczeń.

Unstructured (us)

Każda macierz kowariancyjna może być reprezentowana przez wysycony (saturated) zestaw $k=m(m+1)/2$ parametrów

Compound symmetry (symetria złożona)(cs) i (csh)

Struktury kowariancji symetrii złożonej zakładają stałą korelację między różnymi punktami czasowymi: $\rho_{ij} = \rho$, ρ -jest tutaj pojedynczym parametrem korelacji.

Założenie stałej wariancji w równaniu przestrzeni stanu daje jednorodną strukturę kowariancji symetrii złożonej (cs) z jedynie dwoma parametrami ($\theta = (\sigma, \rho)$), w przeciwnym razie otrzymujemy niejednorodną strukturę kowariancji(csh) symetrii złożonej z $k=m+1$ parametrami $\theta = (\sigma_1, ..., \sigma_m, \rho)$.

$$P = \begin{bmatrix} 1 & \rho & \cdots & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix}, \quad D = \begin{bmatrix} \sigma & 0 & \cdots & \cdots & 0 \\ 0 & \sigma & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma \end{bmatrix} \text{(cs)}, \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_m \end{bmatrix} \text{(csh)}$$

Model efektów losowych- model *II* komponentów wariancyjnych

Dotychczas rozważane modele ANOVA to tzw. **modele efektów stałych** albo **modele I**. W modelu efektów stałych poziomy czynników są z góry ustalone i interesują nas hipotezy liniowe dotyczące tylko tych rozważanych poziomów. Jeżeli z pewnych względów traktujemy poziomy czynnika nie

jako z góry ustalone ale jako wylosowane z pewnej hipotetycznej populacji poziomów to taki model nazywamy modelem efektów losowych (model II)

Model jednoczynnikowej ANOVA - wariant efektów losowych - z k poziomami rozważanego czynnika może być tak jak w przypadku modelu efektów stałych zapisany w postaci

$$X_{ij}=m+\alpha_i+\varepsilon_{ij}, \qquad i=1,...,k;\ j=1,...,n_i$$

gdzie

m oznacza średnią ogólną (w połączonych grupach)

α_i jest obserwacją zmiennej losowej odzwierciedlającą wpływ i tego poziomu rozważanego czynnika wylosowanego z pewnej populacji poziomów. Zakładamy, że

- α_i ma rozkład $N(0,\sigma_\alpha^2)$
- α_i $i=1,...,k$ są niezależne
- ε_{ij} składnik losowy (niezależne zmienne losowe o rozkładzie $N(0,\sigma^2)$)
- zmienne α_i oraz ε_{ij} są niezależne.

Testowana jest hipoteza (**nie jest to hipoteza liniowa !**)

H_0: $\sigma_\alpha^2=0$ przeciwko alternatywie

H_1: $\sigma_\alpha^2\neq 0$.

Dla przypomnienia **model efektów stałych** w tym przypadku jest postaci

$$X_{ij}=m+\alpha_i+\varepsilon_{ij}, \qquad i=1,...,k;\ j=1,...,n_i$$

ε_{ij} iid $N(0,\sigma^2)$

$$\sum_i^k \alpha_i = 0$$

Różnice

- W modelu efektów losowych nie ma sigma ograniczeń, gdyż poziomy są losowe i nie możemy ich kontrolować
- W modelu efektów stałych obserwacje X_{ij} są niezależne i $X_{ij}\sim N(m+\alpha_i,\sigma^2)$ a w modelu efektów losowych $E(X_{ij})=m$- wszystkie obserwacje mają wspólną wartość oczekiwaną

  $V(X_{ij})=\sigma_a^2+\sigma^2$

  W modelu efektów losowych obserwacje z tej samej klasy są skorelowane

$$\rho(X_{ij},X_{ik})=\frac{Cov(X_{ij},X_{ik})}{\sqrt{V(X_{ij})\,V(X_{ik})}}=\frac{E(X_{ij}-m)(X_{ik}-m)}{\sigma_a^2+\sigma^2}=\frac{E(a_i+\varepsilon_{ij})(a_i+\varepsilon_{ik})}{\sigma_a^2+\sigma^2}=\frac{\sigma_a^2}{\sigma_a^2+\sigma^2}$$

Współczynnik korelacji pomiędzy zmiennymi z tej samej klasy (grupy) nosi nazwę współczynnika korelacji wewnątrzklasowej.

W modelu II efektów losowych wektor obserwacji ma więc następujący rozkład

$$\mathbf{X}\sim N\left(\mathbf{1}m,\begin{bmatrix}\mathbf{V_1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{V_k}\end{bmatrix}\right), \text{ gdzie } V_i=\begin{bmatrix}\sigma_a^2+\sigma^2 & \cdots & \sigma_a^2\\ \cdots & \cdots & \cdots\\ \sigma_a^2 & \cdots & \sigma_a^2+\sigma^2\end{bmatrix}$$

Autoregresja (ar1) i (ar1h)

Załóżmy, że reszty w modelu liniowym generowane są przez proces autoregresji $\varepsilon_t=\rho\varepsilon_{t-1}+u_t$, t=1,…,n, gdzie u_t (t=…,-2,-1,0,1,2,…) są niezależnymi zmiennymi losowymi o jednakowych rozkładach, dla których $E(u_t)=0$ i $V(u_t)=E(u_t^2)=\sigma_u^2$. Wówczas składniki losowe ε_t tworzą próbę pochodzącą ze stacjonarnego procesu stochastycznego o parametrach

$$E(\varepsilon_t)=0,\qquad V(\varepsilon_t)=E(\varepsilon_t^2)=\sigma^2,\quad \operatorname{cov}(\varepsilon_t,\varepsilon_s)=\sigma^2\rho^{|t-s|} \text{ gdzie } \sigma^2=\tfrac{\sigma_u^2}{1-\rho^2}.$$

Czyli

$$P=\begin{bmatrix}1 & \rho & \rho^2 & \cdots & \rho^{n-1}\\ \rho & 1 & \rho & \cdots & \rho^{n-2}\\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1\end{bmatrix},\ D=\begin{bmatrix}\sigma & 0 & \cdots & \cdots & 0\\ 0 & \sigma & 0 & \cdots & 0\\ \vdots & & \ddots & & \vdots\\ \vdots & & & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & \sigma\end{bmatrix} \text{ (ar1), gdzie } \sigma^2=\frac{\sigma_u^2}{1-\rho^2}$$

$$D=\begin{bmatrix}\sigma_1 & 0 & \cdots & \cdots & 0\\ 0 & \sigma_2 & 0 & \cdots & 0\\ \vdots & & \ddots & & \vdots\\ \vdots & & & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & \sigma_m\end{bmatrix} \text{ (ar1h), .}$$

Opis pozostałych struktur kowariancyjnych dostępnych w R można znaleźć

https://cran.r-project.org/web/packages/mmrm/vignettes/covariance.html

i w przeglądowym artykule

https://intapi.sciendo.com/pdf/10.2478/bile-2022-0010

W pierwszej z powyższych pozycji można znaleźć informację o algorytmie optymalizacyjnym wykorzystującym procedurę profilowania

Funkcja gls() z biblioteki {nlme} w R realizująca UMNK

This function fits a linear model using generalized least squares. The errors are allowed to be correlated and/or have unequal variances.

Usage

```
gls(model, data, correlation, weights, subset, method, na.action,
    control, verbose)
## S3 method for class 'gls'
update(object, model., ..., evaluate = TRUE)
```

Arguments

object	an object inheriting from class "gls", representing a generalized least squares fitted linear model.
model	a two-sided linear formula object describing the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right.
model.	Changes to the model – see update.formula for details.
data	an optional data frame containing the variables named in model, correlation, weights, and subset. By default the variables are taken from the environment from which gls is called.
correlation	an optional corStruct object describing the within-group correlation structure. See the documentation of corClasses for a description of the available corStruct classes. If a grouping variable is to be used, it must be specified in the form argument to the corStruct constructor. Defaults to NULL, corresponding to uncorrelated errors.
weights	an optional varFunc object or one-sided formula describing the within-group heteroscedasticity structure. If given as a formula, it is used as the argument to varFixed, corresponding to fixed variance weights. See the documentation on varClasses for a description of the available varFunc classes. Defaults to NULL, corresponding to homoscedastic errors.
subset	an optional expression indicating which subset of the rows of data should be used in the fit. This can be a logical vector, or a numeric vector indicating which observation numbers are to be included, or a character vector of the row names to be included. All observations are included by default.
method	a character string. If "REML" the model is fit by maximizing the restricted log-likelihood. If "ML" the log-likelihood is maximized. Defaults to "REML".
na.action	a function that indicates what should happen when the data contain NAs. The default action (na.fail) causes gls to print an error message and terminate if there are any incomplete observations.
control	a list of control values for the estimation algorithm to replace the default values returned by the function glsControl. Defaults to an empty list.
verbose	an optional logical value. If TRUE information on the evolution of the iterative algorithm is printed. Default is FALSE.
...	some methods for this generic require additional arguments. None are used in this method.
evaluate	If TRUE evaluate the new call else return the call.

Możliwe struktury korelacyjne dostępne w użyciu w funkcji gls()

corClasses {nlme} R Documentation

Correlation Structure Classes

Description

Standard classes of correlation structures (corStruct) available in the nlme package.

Value

Available standard classes:

corAR1	autoregressive process of order 1.
corARMA	autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.
corCAR1	continuous autoregressive process (AR(1) process for a continuous time covariate).
corCompSymm	compound symmetry structure corresponding to a constant correlation.

corExp	exponential spatial correlation.
corGaus	Gaussian spatial correlation.
corLin	linear spatial correlation.
corRatio	Rational quadratics spatial correlation.
corSpher	spherical spatial correlation.
corSymm	general correlation matrix, with no additional structure.

Propozycja tematu na zaliczenie.

Przykłady użycia funkcji gls() do estymacji modeli liniowych z różnymi strukturami kowariancyjnymi

Uwagi do metody REML estymacji komponentów wariancyjnych.

Rozważmy model $(\mathbf{Y}, \mathbf{X\beta}, \mathbf{\Sigma})$ tzn. $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$. Idea metody REML (Patterson ,Thompson 1971) polega na estymacji komponentów wariancyjnych (czyli $\mathbf{\Sigma}$) metodą największej wiarygodności zastosowanej do transformowanych danych $\mathbf{KY}$ a nie oryginalnych danych $\mathbf{Y}$, gdzie macierz $\mathbf{K}$ jest tak dobrana aby rozkład $\mathbf{KY}$ zależał tylko od komponentów wariancyjnych i nie zależał od $\boldsymbol{\beta}$. Aby to zrobić, poszukujemy macierzy $\mathbf{K}$ takiej że $\mathbf{KX} = \mathbf{0}$. Zakładamy że macierz $\mathbf{K}$ jest pełnego rzędu oraz, że $\mathbf{KY}$ zawiera tak dużo informacji jak to tylko możliwe o komponentach wariancyjnych, więc $\mathbf{K}$ musi zawierać maksymalną możliwą liczbę wierszy. Można pokazać

Twierdzenie. Macierz $\mathbf{K}$ pełnego rzędu, taka, że $\mathbf{KX} = \mathbf{0}$ jest rozmiaru $(n-r)\times n$, gdzie $r = \text{rank}(\mathbf{X})$. Ponadto $\mathbf{K}$ jest postaci $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{H}) = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X})$ gdzie $\mathbf{C}$ jest transformacją pełnego rzędu wierszy macierzy $\mathbf{I} - \mathbf{H}$ (rzutu ortogonalnego na podprzestrzeń ortogonalną do kolumn macierzy **X**).

Dowód. Wiersze $\mathbf{k}_i^{\mathrm{T}}$ macierzy muszą spełniać równania $\mathbf{k}_i^{\mathrm{T}}\mathbf{X} = \mathbf{0}^{\mathrm{T}}$ lub równoważnie $\mathbf{X}^{\mathrm{T}}\mathbf{k}_i = \mathbf{0}$. Z poniższego twierdzenia

Jeżeli układ równań liniowych $\mathbf{Ax} = \mathbf{b}$ jest niesprzeczny, to wszystkie jego rozwiązania dane są wzorem $\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{h}$ gdzie $\mathbf{A}^{-}$ jest dowolną uogólnioną odwrotnością a $\mathbf{h}$ dowolnym wektorem.

zastosowanego dla $\mathbf{b} = \mathbf{0}$ mamy $\mathbf{k}_i = (\mathbf{I} - (\mathbf{X}^{\mathrm{T}})^{-}\mathbf{X}^{\mathrm{T}})\mathbf{c}$ dla dowolnego wektora $\mathbf{c}$ wymiaru $n\times 1$.

Wobec tego $\mathbf{k}_i^{\mathrm{T}} = \mathbf{c}^{\mathrm{T}}(\mathbf{I} - (\mathbf{X}^{\mathrm{T}})^{-}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{c}^{\mathrm{T}}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-})$ (bo $(\mathbf{A}^{-})^{\mathrm{T}} = (\mathbf{A}^{\mathrm{T}})^{-}$).

Ale $\text{rank}(\mathbf{X}\mathbf{X}^{-}) = \text{rank}(\mathbf{X}) = r$. Ponadto $\mathbf{I} - \mathbf{X}\mathbf{X}^{-}$ jest idempotentna, więc $\text{rank}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-}) = \text{tr}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{X}\mathbf{X}^{-}) = n - r$. Stąd wynika, że istnieje $n-r$ liniowo niezależnych wektorów $\mathbf{k}_i$, spełniających $\mathbf{X}^{\mathrm{T}}\mathbf{k}_i = \mathbf{0}$ i dlatego maksymalna liczba wierszy macierzy $\mathbf{K}$ jest równa $n-r$. Wobec powyższego $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-})$ dla pewnej macierzy $\mathbf{C}$ wymiaru

$(n-r)\times n$ pełnego rzędu ,która specyfikuje liniowo niezależne kombinacje wierszy macierzy $\mathbf{I}-\mathbf{X}\mathbf{X}^{-}$. Przyjmując $\mathbf{X}^{-}=(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}$ (zob. uzupełnienie macierzy) możemy przyjąć $\mathbf{K}=\mathbf{C}(\mathbf{I}-\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X})$.

Dla modelu $\mathbf{Y}\sim N(\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma}))$ mamy $\mathbf{K}\mathbf{Y}\sim N_{n-r}(\mathbf{0},\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})$.

Rozważmy dwa specjalne przypadki.

Jeżeli próba pochodzi z ekonomicznych szeregów czasowych, to naturalnym założeniem będzie przyjęcie, że składniki losowe $\varepsilon_1,...,\varepsilon_n$ pochodzą ze stacjonarnego procesu stochastycznego. Chociaż wszystkie ε_t mają jednakowe rozkłady, to jednak nie są wzajemnie niezależne , bo są skorelowane . Załóżmy że w klasycznym modelu regresji zastąpimy założenie $E(\boldsymbol{\varepsilon})=\mathbf{0}, \mathrm{cov}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}$ założeniem

$$\varepsilon_t=\rho\varepsilon_{t-1}+u_t,\ t=1,\ldots,n,$$

gdzie u_t (t=…,-2,-1,0,1,2,…) są niezależnymi zmiennymi losowymi o jednakowych rozkładach, dla których $E(u_t)=0$ i $V(u_t)=E(u_t^2)=\sigma_u^2$. Wówczas składniki losowe ε_t tworzą próbę pochodzącą ze stacjonarnego procesu stochastycznego o parametrach

$$E(\varepsilon_t)=0,\qquad V(\varepsilon_t)=E(\varepsilon_t^2)=\sigma^2,\quad \mathrm{cov}(\varepsilon_t,\varepsilon_s)=\sigma^2\rho^{|t-s|}\ \text{gdzie}\ \sigma^2=\tfrac{\sigma_u^2}{1-\rho^2}.$$

Uzasadnienie: zapisując równania $\varepsilon_t=\rho\varepsilon_{t-1}+u_t$, t=1,…,n, w postaci operatorowej z operatorem B (przesunięcia w tył tzn. $B\varepsilon_t=\varepsilon_{t-1}$) otrzymujemy

$$(1-\rho B)\varepsilon_t=u_t,\ \text{stąd}\quad \varepsilon_t=(1-\rho B)^{-1}u_t=\sum_{k=0}^{\infty}\rho^k B^k u_t=\sum_{k=0}^{\infty}\rho^k u_{t-k}$$

Stąd otrzymujemy $E(\varepsilon_t)=E(\sum_{k=0}^{\infty}\rho^k u_{t-k})=\sum_{k=0}^{\infty}\rho^k E(u_{t-k})=0$

$$\mathrm{cov}(\varepsilon_t,\varepsilon_s)=E\left((\sum_{i=0}^{\infty}\rho^i u_{t-i})(\sum_{j=0}^{\infty}\rho^j u_{s-j})\right)=\sum_{i,j=0}^{\infty}\rho^{i+j}E(u_{t-i}u_{s-j})$$

Stąd $V(\varepsilon_t)=E(\varepsilon_t^2)=E\left((\sum_{i=0}^{\infty}\rho^i u_{t-i})(\sum_{j=0}^{\infty}\rho^j u_{t-j})\right)=\sum_{i,j=0}^{\infty}\rho^{i+j}E(u_{t-i}u_{t-j})=\sum_{i=0}^{\infty}\rho^{2i}\sigma_u^2=\dfrac{\sigma_u^2}{1-\rho^2}$

$$\mathrm{cov}(\varepsilon_t\varepsilon_{t+k})=E\left((\sum_{i=0}^{\infty}\rho^i u_{t-i})(\sum_{j=0}^{\infty}\rho^j u_{t+k-j})\right)=\sum_{i,j=0}^{\infty}\rho^{i+j}E(u_{t-i}u_{t+k-j})=\sum_{i=0}^{\infty}\rho^{2i+k}\sigma_u^2=\rho^k\frac{\sigma_u^2}{1-\rho^2}$$

Macierz kowariancyjna $\mathrm{cov}(\boldsymbol{\varepsilon})$ jest więc postaci

$$\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma} = \sigma^2\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}, \text{ gdzie } \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}.$$

Można pokazać, że $\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$

Drugi ważny przypadek to sytuacja, gdzie błędy w modelu liniowym są nieskorelowane ale mają różne wariancje . Jeżeli $\boldsymbol{\Sigma} = diag(\frac{1}{w_1},...,\frac{1}{w_n})$ gdzie wagi $w_1,...,w_n$ są znane to , w tym przypadku $\mathbf{P} = diag(\frac{1}{\sqrt{w_1}},...,\frac{1}{\sqrt{w_n}})$ i $\mathbf{P}^{-1} = diag(\sqrt{w_1},...,\sqrt{w_n})$ i $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon}$. Wektor odpowiedzi oraz kolumny macierzy planu eksperymentu należy poddać procedurze „ważenia” i zastosować klasyczny estymator MNK. W efekcie stosujemy metodę ważonych najmniejszych kwadratów

Błędy są proporcjonalne do predykatora $V(\varepsilon_i) \propto x_i$ wówczas $w_i = \frac{1}{x_i}$.

Jeżeli obserwacje Y_i są średnimi z n_i pomiarów, to $V(Y_i) = V(\varepsilon_i) = \frac{\sigma^2}{n_i}$ i wówczas $w_i = n_i$

Oto przykład z eksperymentu mającego na celu zbadanie oddziaływania pewnych rodzajów cząstek elementarnych podczas zderzenia z celami protonowymi. Eksperyment został zaprojektowany w celu przetestowania pewnych teorii na temat natury silnego oddziaływania. Uważa się, że zmienna przekroju czynnego (crossx) jest liniowo związana z odwrotnością energii (energia - została już odwrócona). Na każdym poziomie pędu wykonano bardzo dużą liczbę obserwacji, aby można było dokładnie oszacować odchylenie standardowe odpowiedzi (sd).

```
#Metoda ważonych najmniejszych kwadratów
library(faraway)
data(strongx) #Faraway PRA_62
strongx
g <- lm(crossx ~ energy, strongx, weights=sd^-2)#model ważonych NK
summary(g)

gu <- lm(crossx ~ energy, strongx)#model MNK
summary(gu)
plot(crossx ~ energy, data=strongx)
abline(g)
abline(gu,lty=2)
```

Przykład plik UMNK.R

Rozważmy model liniowy y_t=0.5t +2+ξ_t , t=1,...,50, gdzie wektor reszt ξ_t jest generowany przez następujące mechanizmy losowe

1. ξ_t = -0.9ξ_{t-1} +ε_t , gdzie +ε_t i.i.d. $N(0,1)$
2. ξ_t = 0.9ξ_{t-1} +ε_t , gdzie +ε_t i.i.d. $N(0,1)$

Wygenerować dane do powyższego modelu i sprawdzić jakość procedury Orcuta. Jakie stąd wynikają wnioski co do stosowalności MNK

Transformacja Orcutta i usuwanie autokorelacji

$$\begin{cases} Y_t = \beta_0 + \beta_1 T_t + \varepsilon_t \\ Y_{t-1} = \beta_0 + \beta_1 T_{t-1} + \varepsilon_{t-1} \;/\cdot \hat{\rho} \end{cases}$$

$$Y_t - \hat{\rho} Y_{t-1} = \beta_0(1-\hat{\rho}) + \beta_1(T_t - \hat{\rho} T_{t-1}) + \varepsilon_t - \hat{\rho}\varepsilon_{t-1}$$

$$Y_t^* = \beta_0^* + \beta_1^* T_t^* + \varepsilon_t^*$$

Wykorzystamy dostępny w R zbiór danych regresyjnych Longley'a, gdzie odpowiedzią jest liczba zatrudnionych Employed a predyktorami m.inn. dochód narodowy brutto GNP, liczba ludności }, Population iwspółczynnik zmiany cen GNP.deflator.

	GNP.deflator	GNP	Unemployed	Armed.Forces	Population	Year	Employed
1947	83.0	234.289	235.6	159.0	107.608	1947	60.323
1948	88.5	259.426	232.5	145.6	108.632	1948	61.122
1949	88.2	258.054	368.2	161.6	109.773	1949	60.171
1950	89.5	284.599	335.1	165.0	110.929	1950	61.187
1951	96.2	328.975	209.9	309.9	112.075	1951	63.221
1952	98.1	346.999	193.2	359.4	113.270	1952	63.639
1953	99.0	365.385	187.0	354.7	115.094	1953	64.989
1954	100.0	363.112	357.8	335.0	116.219	1954	63.761
1955	101.2	397.469	290.4	304.8	117.388	1955	66.019
1956	104.6	419.180	282.2	285.7	118.734	1956	67.857
1957	108.4	442.769	293.6	279.8	120.445	1957	68.169
1958	110.8	444.546	468.1	263.7	121.950	1958	66.513
1959	112.6	482.704	381.3	255.2	123.366	1959	68.655
1960	114.2	502.601	393.1	251.4	125.368	1960	69.564
1961	115.7	518.173	480.6	257.2	127.852	1961	69.331
1962	116.9	554.894	400.7	282.7	130.081	1962	70.551

Dopasujmy model liniowy zależności

```
data(longley) #Faraway PRA str 60
g <- lm(Employed ~ GNP + Population, data=longley)
summary(g,cor=T)
cor(g$res[-1],g$res[-16])# estymujemy współczynnik autokorelacji reszt
x <- model.matrix(g)
Sigma <- diag(16)# tworzymy macierz jednostkową I_16
Sigma <- 0.31041^abs(row(Sigma)-col(Sigma))# tworzymy macierz kowariancyjną dla błędów z procesu AR(1)
Sigi <- solve(Sigma)# tworzymy odwrotność powyższej macierzy
xtxi <- solve(t(x) %*% Sigi %*% x)# wyznaczamy odwrotność macierzu X'(Sigma^-1) X
beta <- xtxi %*% t(x) %*% Sigi %*% longley$Empl # estymatory UMNK
```

```
beta
res <- longley$Empl - x %*% beta #reszty z UMNK
sig <- sqrt(sum(res^2)/g$df)
sqrt(diag(xtxi))*sig # błędy std estymatorów UMNK
cor(res[-1],res[-16]) # nowy współczynnik autokorelacji reszt

sm<-chol(Sigma)
smi<-solve(t(sm))
sx<- smi %*% x
sy<- smi %*%longley$Empl
lm(sy~sx-1)$coef


library(nlme)
#estymacja metodą REML
g_REML <- gls(Employed ~ GNP + Population,
       correlation=corAR1(form= ~Year), data=longley)
summary(g_REML)
intervals(g_REML)
#estymacja metodą ML
g_ML <- gls(Employed ~ GNP + Population,
            correlation=corAR1(form= ~Year),method="ML", data=longley)
summary(g_ML)
intervals(g_ML)


#Metoda ważonych najmniejszych kwadratów
library(faraway)
data(strongx) #Faraway PRA str 62
strongx
g <- lm(crossx ~ energy, strongx, weights=sd^-2)#model ważonych NK
summary(g)

gu <- lm(crossx ~ energy, strongx)#model MNK
summary(gu)
plot(crossx ~ energy, data=strongx)
abline(g)
abline(gu,lty=2)
```