Lecture 9 | 23.04.2024

Linear regression models

with heteroscedasetic errors

Normal linear model

- **Assumptions:** random sample $(Y_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$ such that $Y_i|\boldsymbol{X}_i \sim N(\boldsymbol{X}_i^\top \beta, \sigma^2)$
- **Inference:** confidence intervals for β_j, confidence regions for β and linear combinations of the form $\mathbb{L}\beta$ (corresponding statistical tests)

Parameter estimates $\widehat{\beta}$ (constructed in terms of LSE or MLE) are BLUE and the follow the normal distribution with the corresponding mean vector and the variance-covariance matrix

Statistical inference is exact, based on the normal distribution (if the variance parameter is known) or the Student's t-distribution or Fisher's F-distribution respectively

Assumptions for a model without normality

Assumptions (A1)

- random sample $(Y_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$
- mean specification $E[Y_i|\boldsymbol{X}_i] = \boldsymbol{X}_i^\top \beta$, respectively $E[\boldsymbol{Y}|\mathbb{X}] = \mathbb{X}\beta$
- thus, for errors $\varepsilon_i = Y_i - \boldsymbol{X}_i^\top \beta$ we have $E[\varepsilon_i|\boldsymbol{X}_i] = E[Y_i - \boldsymbol{X}_i^\top \beta|\boldsymbol{X}_i] = 0$ and $Var(\varepsilon_i|\boldsymbol{X}_i) = Var[Y_i - \boldsymbol{X}_i^\top \beta|\boldsymbol{X}_i] = Var[Y_i|\boldsymbol{X}_i] = \sigma^2(\boldsymbol{X}_i)$
- and for unconditional expectations, $E[\varepsilon_i] = E[E[\varepsilon_i|\boldsymbol{X}_i]] = 0$ and $Var(\varepsilon_i) = Var(E[\varepsilon_i|\boldsymbol{X}_i]) + E[Var(\varepsilon_i|\boldsymbol{X}_i)] = Var(0) + E[\sigma^2(\boldsymbol{X}_i)] = E[\sigma^2(\boldsymbol{X}_i)]$

Assumptions (A2)

- $E|X_j X_k| < \infty$ for $j, k \in \{1, \ldots, p\}$
- $E\left(\boldsymbol{X}\boldsymbol{X}^\top\right) = \mathbb{W} \in \mathbb{R}^{p \times p}$ is a positive definite matrix
- $\mathbb{V} = \mathbb{W}^{-1}$

Assumptions (A3a/A3b)

- **Homoscedastic model)**
  $\sigma^2(\boldsymbol{X}) = Var(Y|\boldsymbol{X}) = \sigma^2 > 0$
- **Heteroscedastic model**
  $\sigma^2(\boldsymbol{X}) = Var(Y|\boldsymbol{X})$ such that $E[\sigma^2(\boldsymbol{X})] < \infty$ and moreover, it also holds that $E[\sigma^2(\boldsymbol{X})X_j X_k] < \infty$ for $j, k \in \{1, \ldots, p\}$

Inference under (A1), (A2), and (A3a)

Linear model without normality assumption (homoscedastic errors)

- Parameter estimates $\widehat{\beta}_n$ (constructed in terms of LSE) are BLUE, they are consistent (convergence in probability for $n \to \infty$) and they follow asymptotically the normal distribution with the corresponding mean vector and the variance-covariance matrix $\sigma^2\mathbb{V}$

- Statistical inference is approximate (asymptotical) based on the asymptotic normal distribution (central limit theorem)
(inference in terms of statistical tests and confidence intervals)

Heteroscedasticity in a linear model

Basically, there are two different model frameworks where we need to deal with heteroscedasticity issues...

- **General linear model**
  the model with heteroscedastic errors but the variance structure of the observations (errors respectively) is apriori known (e.g., it is determined by the design of the experiment)

- **Linear model with hetereoscedastic errors**
  in this case the variance structure of the error terms (observations respectively) is fully unknown and it needs to be estimated if some statistical inference is of any interest

General linear model

- random sample $(Y_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$
- mean specification $E[\boldsymbol{Y}|\mathbb{X}] = \mathbb{X}\beta$, for $\beta \in \mathbb{R}^p$
- variance specification $Var[\boldsymbol{Y}|\mathbb{X}] = \sigma^2\mathbb{W}^{-1}$, for some known matrix $\mathbb{W} \in \mathbb{R}^{n\times n}$ (positive definite)
- generally, the normal distribution is not assumed, thus

$$\boldsymbol{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2\mathbb{W}^{-1}),$$

but the normal distribution can be postulated and the results hold analogously as in the normal linear model

General linear model

- random sample $(Y_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$
- mean specification $E[\boldsymbol{Y}|\mathbb{X}] = \mathbb{X}\beta$, for $\beta \in \mathbb{R}^p$
- variance specification $Var[\boldsymbol{Y}|\mathbb{X}] = \sigma^2\mathbb{W}^{-1}$, for some known matrix $\mathbb{W} \in \mathbb{R}^{n\times n}$ (positive definite)
- generally, the normal distribution is not assumed, thus

$$\boldsymbol{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2\mathbb{W}^{-1}),$$

but the normal distribution can be postulated and the results hold analogously as in the normal linear model

Example

Consider a linear regression model, where the dependent variables Y_i for $i = 1, \ldots, n$ represent some averages across $w_i \in \mathbb{N}$ independent subjects, where for each subject we assume the same variance (a homoscedastic model for the subjects)

General least squares

Consider a general linear model $\boldsymbol{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2\mathbb{W}^{-1})$ where $rank(\mathbb{X}) = p < n$ (where $\mathbb{X} \in \mathbb{R}^{n\times p}$) Than the following holds:

- $\widehat{\beta}_G = (\mathbb{X}^\top\mathbb{W}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{W}\boldsymbol{Y}$ is BLUE for $\beta \in \mathbb{R}^p$
- $\widehat{\mu} = \widehat{\boldsymbol{Y}} = \mathbb{X}\widehat{\beta}_G$ is BLUE for $\mu = E[\boldsymbol{Y}|\mathbb{X}]$
- for $\boldsymbol{l} \in \mathbb{R}^p$, where $\boldsymbol{l} \neq \boldsymbol{0}$, $\widehat{\theta} = \boldsymbol{l}^\top\widehat{\beta}_G$ is BLUE for $\theta = \boldsymbol{l}^\top\beta$
- $SSe_G = \|\mathbb{W}^{1/2}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})\|_2^2 = (\boldsymbol{Y} - \widehat{\boldsymbol{Y}})^\top\mathbb{W}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})$ is the generalized residual sum of squares
- $MSe_G = \frac{1}{n-p}\|\mathbb{W}^{1/2}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})\|_2^2$ is an unbiased estimate of $\sigma^2 > 0$

General least squares

Consider a general linear model $\boldsymbol{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2\mathbb{W}^{-1})$ where $rank(\mathbb{X}) = p < n$ (where $\mathbb{X} \in \mathbb{R}^{n\times p}$) Than the following holds:

- $\widehat{\beta}_G = (\mathbb{X}^\top\mathbb{W}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{W}\boldsymbol{Y}$ is BLUE for $\beta \in \mathbb{R}^p$
- $\widehat{\mu} = \widehat{\boldsymbol{Y}} = \mathbb{X}\widehat{\beta}_G$ is BLUE for $\boldsymbol{\mu} = E[\boldsymbol{Y}|\mathbb{X}]$
- for $\boldsymbol{l} \in \mathbb{R}^p$, where $\boldsymbol{l} \neq \boldsymbol{0}$, $\widehat{\theta} = \boldsymbol{l}^\top\widehat{\beta}_G$ is BLUE for $\theta = \boldsymbol{l}^\top\beta$
- $SSe_G = \|\mathbb{W}^{1/2}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})\|_2^2 = (\boldsymbol{Y} - \widehat{\boldsymbol{Y}})^\top\mathbb{W}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})$ is the generalized residual sum of squares
- $MSe_G = \frac{1}{n-p}\|\mathbb{W}^{1/2}(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})\|_2^2$ is an unbiased estimate of $\sigma^2 > 0$

If, additionaly, $\boldsymbol{Y}|\mathbb{X} \sim N(\mathbb{X}\beta, \sigma^2\mathbb{W}^{-1})$ then the estimates follow the corresponding normal distribution and

$$\frac{MSe_G(n-p)}{\sigma^2} = \frac{SSe_G}{\sigma^2} \sim \chi^2_{n-p}$$

and SSe and $\widehat{\boldsymbol{Y}}$ are conditionally, given $\mathbb{X}$, mutually independent

General linear model – utilization

- the general linear model is typically used with partially aggregated data—mostly in a way, that instead of raw observations we observe independent averages over specific classes (that we can control for with the set of the regressor variables)
- if the estimation of the mean structure if of the interest only, the aggregated data can be also replicated and the correponding mean estiamates will be analogous
- however, if there is also some interest in the variance estimation (e.g., there is a need to perform some statistical inference) than the model based on replicated data will fail (the variance estimates are artificially underestimated)

More general situations...

- General least squares represent a class of linear models for heteroscedastic data, however, with the known heteroscedastic structure—the matrix $\mathbb{W}$ is known in advance, e.g., from the experiment
- More general scenario involves situations where heteroscedastic data have some unknown variance structure (which needs to be estimated)

More general situations...

- General least squares represent a class of linear models for heteroscedastic data, however, with the known heteroscedastic structure—the matrix $\mathbb{W}$ is known in advance, e.g., from the experiment
- More general scenario involves situations where heteroscedastic data have some unknown variance structure (which needs to be estimated)

- Recall Assumption (A3) that specified the following conditions:
  - **Heteroscedastic model**
    $\sigma^2(\boldsymbol{X}) = Var(Y|\boldsymbol{X})$ such that $E[\sigma^2(\boldsymbol{X})] < \infty$ and moreover, it also holds that $E[\sigma^2(\boldsymbol{X})X_jX_k] < \infty$ for $j, k \in \{1, \ldots, p\}$

More general situations...

- General least squares represent a class of linear models for heteroscedastic data, however, with the known heteroscedastic structure—the matrix $\mathbb{W}$ is known in advance, e.g., from the experiment
- More general scenario involves situations where heteroscedastic data have some unknown variance structure (which needs to be estimated)
- Recall Assumption (A3) that specified the following conditions:
  - **Heteroscedastic model**
    $\sigma^2(\boldsymbol{X}) = Var(Y|\boldsymbol{X})$ such that $E[\sigma^2(\boldsymbol{X})] < \infty$ and moreover, it also holds that $E[\sigma^2(\boldsymbol{X})X_jX_k] < \infty$ for $j, k \in \{1, \ldots, p\}$
- The assumption above implies, that the matrix $\mathbb{W}^\star = E[\sigma^2(\boldsymbol{X})\boldsymbol{X}\boldsymbol{X}^\top]$ is a real matrix with all elements being finite
- Thus, under the heteroscedastic model, we have $E[Y_i|\boldsymbol{X}_i] = \boldsymbol{X}_i^\top\beta$ and $Var[Y_i|\boldsymbol{X}_i] = Var[\varepsilon_i|\boldsymbol{X}_i] = \sigma^2(\boldsymbol{X}_i)$

Consistency of the LSE estimates

The underlying model can be either assumed within the normal model framework or, alternatively, no normality is needed

- Again, we are interested in the following parameters:
  - $\boldsymbol{\beta} \in \mathbb{R}^p$
  - $\sigma^2 > 0$
  - $\theta = \boldsymbol{l}^\top \boldsymbol{\beta} \in \mathbb{R}$, for some nonzero vector $\boldsymbol{l} \in \mathbb{R}^p$
  - $\Theta = \mathbb{L}\boldsymbol{\beta} \in \mathbb{R}^m$, for some matrix $\mathbb{L} \in \mathbb{R}^{m \times p}$ with linearly independent rows

Consistency of the LSE estimates

The underlying model can be either assumed within the normal model framework or, alternatively, no normality is needed

- Again, we are interested in the following parameters:
  - $\boldsymbol{\beta} \in \mathbb{R}^p$
  - $\sigma^2 > 0$
  - $\theta = \boldsymbol{l}^\top \boldsymbol{\beta} \in \mathbb{R}$, for some nonzero vector $\boldsymbol{l} \in \mathbb{R}^p$
  - $\Theta = \mathbb{L}\boldsymbol{\beta} \in \mathbb{R}^m$, for some matrix $\mathbb{L} \in \mathbb{R}^{m \times p}$ with linearly independent rows

- The corresponding estmates are defined straightforwardly and it holds (under (A1), (A2), and (A3a/A3b)) that
  - $\widehat{\boldsymbol{\beta}}_n \longrightarrow \boldsymbol{\beta}$ a.s. (in P), for $n \to \infty$
  - $\widehat{\theta}_n = \boldsymbol{l}^\top \widehat{\boldsymbol{\beta}}_n \longrightarrow \theta$ a.s. (in P), for $n \to \infty$
  - $\widehat{\Theta}_n = \mathbb{L}\widehat{\boldsymbol{\beta}}_n \longrightarrow \Theta$, a.s. (in P), for $n \to \infty$

Assymptotic normality under heteroscedasticity

Under the assumptions stated in (A1), (A2), and (A3b) and, additionally, for $E[\varepsilon^2 X_j X_k] < \infty$ for $j, k = 1, \ldots, p$ the following holds:

Assymptotic normality under heteroscedasticity

Under the assumptions stated in (A1), (A2), and (A3b) and, additionally, for $E[\varepsilon^2 X_j X_k] < \infty$ for $j, k = 1, \ldots, p$ the following holds:

- $\sqrt{n}(\widehat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} N_p(\beta, \mathbb{V}\mathbb{W}^\star\mathbb{V})$ for $n \to \infty$
- $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{l}^\top \mathbb{V}\mathbb{W}^\star\mathbb{V}\boldsymbol{l})$, as $n \to \infty$
- $\sqrt{n}(\widehat{\Theta}_n - \Theta) \xrightarrow{\mathcal{D}} N_m(\mathbf{0}, \mathbb{L}\mathbb{V}\mathbb{W}^\star\mathbb{V}\mathbb{L}^\top)$, as $n \to \infty$

where $\mathbb{V} = \left[E(\boldsymbol{X}\boldsymbol{X}^\top)\right]^{-1}$ and $\mathbb{W}^\star = E[\sigma^2(\boldsymbol{X})\boldsymbol{X}\boldsymbol{X}^\top]$

Note, that $Var(\boldsymbol{X}\varepsilon) = E[\sigma^2(\boldsymbol{X})\boldsymbol{X}\boldsymbol{X}^\top]$ which equals to $\sigma^2 E[\boldsymbol{X}\boldsymbol{X}^\top] = \sigma^2\mathbb{W}$ under homoscedasticity (A3a) and it equals to $\mathbb{W}^\star$ under heteroscedasticity (A3b)

Sandwich estimate of the variance

Consider the assumptions in (A1), (A2), and (A3b). Let, moreover, the following holds

- $E|\varepsilon^2 X_j X_k| < \infty$
- $E|\varepsilon X_j X_k X_s| < \infty$
- $E|X_j X_l X_s X_l| < \infty$

all for $j, k, s, l \in \{1, \ldots, p\}$. Then the following holds:

$$n\mathbb{V}_n \mathbb{W}_n^\star \mathbb{V}_n \xrightarrow{a.s.(P)} \mathbb{V}\mathbb{W}^\star\mathbb{V}, \quad \text{for } n \to \infty$$

where $\mathbb{W}_n^\star = \sum_{i=1}^n U_i^2 \boldsymbol{X}_i \boldsymbol{X}_i^\top = \mathbb{X}_n^\top \boldsymbol{\Omega}_n \mathbb{X}_n$, where $U_i = Y_i - \widehat{Y}_i$ and $\boldsymbol{\Omega}_n = diag(U_1^2, \ldots, U_n^2)$

Sandwich estimate of the covariance matrix

- the estimate for the variance covariance matrix $\mathbb{V}\mathbb{W}^\star\mathbb{V}$ is the so-called **sandwich estimate** of the form

$$\mathbb{V}_n\mathbb{W}_n^\star\mathbb{V}_n = \underbrace{(\mathbb{X}_n^\top\mathbb{X}_n)^{-1}\mathbb{X}_n^\top}_{\textit{bread}} \quad \underbrace{\boldsymbol{\Omega_n}}_{\textit{mean}} \quad \underbrace{\mathbb{X}_n(\mathbb{X}_n^\top\mathbb{X}_n)^{-1}}_{\textit{bread}}$$

  which is a (heteroscedastic) consistent estimate of the variance-covarance of the least squares estimate $\widehat{\beta}_n$

- if we replace the matrix $\boldsymbol{\Omega_n}$ with $\frac{n}{\nu_n}\boldsymbol{\Omega}_n$ for some sequence $\{\nu_n\}_n$ such that $n/\nu_n \to 1$ as $n \to \infty$ the convergence still holds and ν_n is called the **degrees of freedom of the sandwich** estimate
- different options are used in the literature to define the sequence $\{\nu_n\}_n$ (White (1980); MacKinnon and White (1985); etc.)

Asymptotic inference under heteroscedasticity

- for a (consistent) sandwich estimate $\mathbb{V}_n^{HC} = (\mathbb{X}_n^\top \mathbb{X}_n)^{-1} \mathbb{X}_n^\top \boldsymbol{\Omega_n} \mathbb{X}_n (\mathbb{X}_n^\top \mathbb{X}_n)^{-1}$ of the covariance matrix of the LSE $\widehat{\beta}_n$ we can define
  - $T_n = \dfrac{\boldsymbol{l}^\top \widehat{\beta} - \boldsymbol{l}^\top \beta}{\sqrt{\boldsymbol{l}^\top \mathbb{V}_n^{HC} \boldsymbol{l}}}$
  - $Q_n = \dfrac{(\mathbb{L}\widehat{\beta}_n - \mathbb{L}\beta)^\top \left(\mathbb{L}\mathbb{V}_n^{HC}\mathbb{L}^\top\right)^{-1} (\mathbb{L}\widehat{\beta}_n - \mathbb{L}\beta)}{m}$

Asymptotic inference under heteroscedasticity

- for a (consistent) sandwich estimate $\mathbb{V}_n^{HC} = (\mathbb{X}_n^\top \mathbb{X}_n)^{-1} \mathbb{X}_n^\top \boldsymbol{\Omega_n} \mathbb{X}_n (\mathbb{X}_n^\top \mathbb{X}_n)^{-1}$ of the covariance matrix of the LSE $\widehat{\beta}_n$ we can define
  - $T_n = \frac{\boldsymbol{l}^\top \widehat{\beta} - \boldsymbol{l}^\top \beta}{\sqrt{\boldsymbol{l}^\top \mathbb{V}_n^{HC} \boldsymbol{l}}}$
  - $Q_n = \frac{(\mathbb{L}\widehat{\beta}_n - \mathbb{L}\beta)^\top \left(\mathbb{L}\mathbb{V}_n^{HC}\mathbb{L}^\top\right)^{-1} (\mathbb{L}\widehat{\beta}_n - \mathbb{L}\beta)}{m}$
- Statistic T_n follows (asymptotically) the normal distribution $N(0,1)$ and the statistic mQ_n follows (again asymptotically) the χ^2 distribution with $m = rank(\mathbb{L})$ degrees of freedom (for $n \to \infty$)
- Note, that the results are analogous to those obtained for the homoscedastic case where $MSe(\mathbb{X}^\top \mathbb{X})^{-1}$ is replaced by the sandwich estimate $\mathbb{V}_n^{HC}$

Summary

- **Linear regression models...**
  - Normal linear model with homoscedastic errors
  - Linear model without normality assumptions (A3a/A3b)
  - General linear model (with and without the normality assumption)
- **Consistent LSE/MLE estimates**
  - consistent estimates of the mean and variance parameters
  - the mean parameter estimates are normally distributed (normal model)
  - the mean estimates are asymptotically normal (model without normality)
  - consistent estimates of the variance parameter/parameters
- **Statistical inference**
  - primarily about the mean parameters and their linear combinations
  - exact and approximate (asymptotic) confidence intervals (regions)
  - statistical tests