

1 LMVAR: a linear model with heteroscedasticity

This vignette describes in more detail the mathematical aspects of the model with which the `lmvar` package is concerned. A short description can be found in the vignette 'Intro' of this package.

Assume that a stochastic vector $Y \in \mathbb{R}^n$ has a multivariate normal distribution as

$$Y \sim \mathcal{N}_n(\mu^*, \Sigma^*) \quad (1)$$

in which $\mu^* \in \mathbb{R}^n$ is the expected value and $\Sigma^* \in \mathbb{R}^{n,n}$ a diagonal covariance matrix

$$\Sigma_{ij}^* = \begin{cases} 0 & i \neq j \\ (\sigma_i^*)^2 & i = j. \end{cases} \quad (2)$$

Assume that the vector of expectation values μ^* is linearly dependent on the values of the covariates in a model matrix X_μ :

$$\mu^* = X_\mu \beta_\mu^* \quad (3)$$

with $X_\mu \in \mathbb{R}^{n,k_\mu}$ and $\beta_\mu^* \in \mathbb{R}^{k_\mu}$.

Similarly, assume that the vector $\sigma^* = (\sigma_1^*, \dots, \sigma_n^*)$ depends on the covariates in a model matrix X_σ as

$$\log \sigma^* = X_\sigma \beta_\sigma^* \quad (4)$$

where $\log \sigma^* = (\log \sigma_1^*, \dots, \log \sigma_n^*)$, $X_\sigma \in \mathbb{R}^{n,k_\sigma}$ and $\beta_\sigma^* \in \mathbb{R}^{k_\sigma}$. The logarithm is taken to be the 'natural logarithm', i.e., with base e .

We assume $n \geq k_\mu + k_\sigma$ to avoid having an overdetermined system when we calculate estimators for β_μ^* and β_σ^* , as explained in the next section.

If we take X_σ a $n \times 1$ matrix in which each element is equal to 1, we have the standard linear model.

The parameter vector β_μ^* is defined uniquely only if X_μ is full-rank. If not, the space \mathbb{R}^{k_μ} can be split into subspaces such that there is a uniquely defined β_μ^* in each subspace. The way `lmvar` treats this is as follows. If the user-supplied X_μ is not full-rank, `lmvar` removes just enough columns from the matrix to make it full-rank. This amounts to selecting β_μ^* from the subspace in which all vector elements corresponding to the removed columns, are set to zero.

In the same way, if the user-supplied X_σ is not full-rank, just enough columns are removed to make it so. This defines a subspace in which β_σ^* is defined uniquely.

In what follows we assume that X_μ and X_σ are the matrices after the columns have been removed, i.e., they are full-rank matrices. The vector elements that are set to zero, drop out of β_μ^* and β_σ^* and the dimensions k_μ and k_σ are reduced accordingly. These reduced dimensions are returned by the function `dfree` in the `lmvar` package.

2 Maximum-likelihood equations

A vector element Y_i is distributed as

$$Y_i \sim \frac{1}{\sqrt{2\pi\sigma_i^*}} \exp\left(-\frac{1}{2} \left(\frac{Y_i - \mu_i^*}{\sigma_i^*}\right)^2\right). \quad (5)$$

The logarithm of the likelihood \mathcal{L} is defined as

$$\log \mathcal{L}(\beta_\mu, \beta_\sigma) = -\frac{n}{2} \log(2\pi) - \sum_{k=1}^n \left(\log \sigma_k + \frac{(y_k - \mu_k)^2}{2\sigma_k^2} \right). \quad (6)$$

for all vectors $\beta_\mu \in \mathbb{R}^{k_\mu}$ and $\beta_\sigma \in \mathbb{R}^{k_\sigma}$ and μ and σ defined as

$$\begin{aligned} \mu &= X_\mu \beta_\mu \\ \log \sigma &= X_\sigma \beta_\sigma. \end{aligned} \quad (7)$$

We are looking for $\hat{\beta}_\mu \in \mathbb{R}^{k_\mu}$ and $\hat{\beta}_\sigma \in \mathbb{R}^{k_\sigma}$ that maximize the log-likelihood:

$$(\hat{\beta}_\mu, \hat{\beta}_\sigma) = \underset{(\beta_\mu, \beta_\sigma) \in \mathbb{R}^{k_\mu} \times \mathbb{R}^{k_\sigma}}{\operatorname{argmax}} \log \mathcal{L}(\beta_\mu, \beta_\sigma). \quad (8)$$

These maximum likelihood estimators are taken to be the estimators of β_μ^* and β_σ^* . We assume that $\hat{\beta}_\mu$ and $\hat{\beta}_\sigma$ thus defined, exist and are unique.

Given $\hat{\beta}_\sigma$, this is true for $\hat{\beta}_\mu$. Namely, given any β_σ , $\log \mathcal{L}$ is maximized by the β_μ which is the solution of

$$\nabla_{\beta_\mu} \log \mathcal{L} = 0 \quad (9)$$

where ∇_{β_μ} stands for the gradient $(\frac{\partial}{\partial \beta_{\mu,1}}, \dots, \frac{\partial}{\partial \beta_{\mu,n}})$.

This solution is

$$\beta_\mu = (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1} y. \quad (10)$$

with $\Sigma \in \mathbb{R}^{n,n}$ defined as in (2) but with β_σ arbitrary:

$$\Sigma_{ij} = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j. \end{cases} \quad (11)$$

Because of our assumption that X_μ is full rank, the inverse of the matrix $X_\mu^T \Sigma^{-1} X_\mu$ can be taken.

It is easy to see that the solution (10) represents a maximum in the log-likelihood. The matrix $H_{\mu\mu}$ of second-order derivatives

$$(H_{\mu\mu})_{ij} = \frac{\partial^2 \log L}{\partial \beta_{\mu i} \partial \beta_{\mu j}} \quad (12)$$

is given by

$$H_{\mu\mu} = -X_\mu^T \Sigma^{-1} X_\mu, \quad (13)$$

which is negative-definite for any β_σ .

Our maximization search can now be carried out in a smaller space:

$$\hat{\beta}_\sigma = \operatorname{argmax}_{\beta_\sigma \in \mathbb{R}^{k_\sigma}} \log \mathcal{L}_P(\beta_\sigma) \quad (14)$$

where \mathcal{L}_P is the so-called profile-likelihood

$$\mathcal{L}_P(\beta_\sigma) = \mathcal{L}(\beta_\mu(\beta_\sigma), \beta_\sigma). \quad (15)$$

with β_μ depending on β_σ as in (10).

To find $\hat{\beta}_\sigma$ from (14), we must solve

$$(\nabla_{\beta_\mu} \log \mathcal{L}) (\nabla_{\beta_\sigma} \beta_\mu) + \nabla_{\beta_\sigma} \log \mathcal{L} = 0 \quad (16)$$

evaluated at $\beta_\mu = \beta_\mu(\beta_\sigma)$, and $(\nabla_{\beta_\sigma} \beta_\mu)$ the matrix

$$(\nabla_{\beta_\sigma} \beta_\mu)_{ij} = \frac{\partial \beta_{\mu i}}{\partial \beta_{\sigma j}}. \quad (17)$$

However, because of (9), the first term in (16) vanishes and we are left to solve

$$\nabla_{\beta_\sigma} \log \mathcal{L} = 0. \quad (18)$$

The derivatives that are the elements of this gradient are given by

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \beta_{\sigma i}} &= \sum_{k=1}^n \left(-(X_\sigma)_{ki} + \frac{(y_k - \mu_k)^2}{\sigma_k^2} (X_\sigma)_{ki} \right) \\ &= \sum_{k=1}^n \left(\frac{(y_k - \mu_k)^2}{\sigma_k^2} - 1 \right) (X_\sigma)_{ki}. \end{aligned} \quad (19)$$

The entire gradient can be written as a matrix-product as

$$\nabla_{\beta_\sigma} \log \mathcal{L} = X_\sigma^T \lambda_\sigma \quad (20)$$

with λ_σ a vector of length n whose elements $\lambda_{\sigma i}$ are

$$\lambda_{\sigma i} = \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 - 1. \quad (21)$$

The maximum-likelihood equations (18) take the form

$$X_\sigma^T \lambda_\sigma = 0. \quad (22)$$

The estimate μ of the expectation value that appears in λ_σ depends on β_σ as

$$\begin{aligned} \mu &= X_\mu \beta_\mu \\ &= X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1} y \\ &= \Sigma^{-1/2} X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1/2} y \end{aligned} \quad (23)$$

where the latter form is the more symmetric, with

$$\left(\Sigma^{-1/2}\right)_{ij} = \begin{cases} 0 & i \neq j \\ \frac{1}{\sigma_i} & i = j. \end{cases} \quad (24)$$

The vector $(y - \mu)/\sigma$, which i -th element is $(y_i - \mu_i)/\sigma_i$, can be written as

$$\frac{y - \mu}{\sigma} = \Sigma^{-1/2} \left[I - \Sigma^{-1/2} X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1/2} \right] y \quad (25)$$

in which $I \in \mathbb{R}^{n,n}$ is the identity matrix.

2.1 Profile-likelihood Hessian

Numerical procedures to solve the maximum-likelihood equations $X_\sigma^T \lambda_\sigma = 0$ involve the calculation of the Hessian H_P of the profile log-likelihood. H_P is the matrix of second-order derivatives of $\log \mathcal{L}_P$:

$$(H_P)_{ij} = \frac{\partial^2 \log \mathcal{L}_P}{\partial \beta_{\sigma j} \partial \beta_{\sigma i}} \quad (26)$$

Differentiation of (19) gives for the second-order derivatives

$$(H_P)_{ij} = -2 \sum_{k=1}^n (X_\sigma^T)_{ik} \frac{y_k - \mu_k}{\sigma_k^2} \left\{ \frac{\partial \mu_k}{\partial \beta_{\sigma j}} + (y_k - \mu_k)(X_\sigma)_{kj} \right\} \quad (27)$$

with $\partial \mu_k / (\partial \beta_{\sigma j})$ the element at row k and column j of the matrix $(\nabla_{\beta_\sigma} \mu)$. Given that $\mu = X_\mu \beta_\mu$ and β_μ is given by (10), the j -th column vector of the matrix is

$$\begin{aligned} \frac{\partial \mu}{\partial \beta_{\sigma j}} &= X_\mu \frac{\partial \beta_\mu}{\partial \beta_{\sigma j}} \\ &= X_\mu \left\{ \frac{\partial (X_\mu^T \Sigma^{-1} X_\mu)^{-1}}{\partial \beta_{\sigma j}} X_\mu^T \Sigma^{-1} + (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} \right\} y \\ &= X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} \left\{ -X_\mu^T \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1} + X_\mu^T \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} \right\} y \\ &= X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} \left\{ -X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \Sigma^{-1} + I \right\} y \\ &= X_\mu (X_\mu^T \Sigma^{-1} X_\mu)^{-1} X_\mu^T \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} (y - \mu) \end{aligned} \quad (28)$$

The matrix $\partial \Sigma^{-1} / (\partial \beta_{\sigma j})$ takes the form

$$\begin{aligned} \frac{\partial \Sigma^{-1}}{\partial \beta_{\sigma j}} &= \sum_{i=1}^n \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta_{\sigma j}} \\ &= -2 \begin{pmatrix} (X_\sigma)_{1j} & & 0 \\ & \ddots & \\ 0 & & (X_\sigma)_{nj} \end{pmatrix} \Sigma^{-1} \end{aligned} \quad (29)$$

The j -th column vector of the matrix is

$$\frac{\partial \mu}{\partial \beta_{\sigma j}} = -2X_{\mu} (X_{\mu}^T \Sigma^{-1} X_{\mu})^{-1} X_{\mu}^T \begin{pmatrix} \frac{y_1 - \mu_1}{\sigma_1^2} (X_{\sigma})_{1j} \\ \vdots \\ \frac{y_n - \mu_n}{\sigma_n^2} (X_{\sigma})_{nj} \end{pmatrix} \quad (30)$$

and the element $(\nabla_{\beta_{\sigma}} \mu)_{kj}$ of the matrix $(\nabla_{\beta_{\sigma}} \mu)$ is given by

$$\frac{\partial \mu_k}{\partial \beta_{\sigma j}} = -2 \sum_{l=1}^n \left(X_{\mu} (X_{\mu}^T \Sigma^{-1} X_{\mu})^{-1} X_{\mu}^T \right)_{kl} \frac{y_l - \mu_l}{\sigma_l^2} (X_{\sigma})_{lj}. \quad (31)$$

If we substitute this result in (27), we obtain for the element at row i and column j of the Hessian:

$$\begin{aligned} (H_P)_{ij} = & 4 \sum_{k,l=1}^n (X_{\sigma}^T)_{ik} \frac{y_k - \mu_k}{\sigma_k^2} \left(X_{\mu} (X_{\mu}^T \Sigma^{-1} X_{\mu})^{-1} X_{\mu}^T \right)_{kl} \frac{y_l - \mu_l}{\sigma_l^2} (X_{\sigma})_{lj} + \\ & - 2 \sum_{k=1}^n (X_{\sigma}^T)_{ik} \left(\frac{y_k - \mu_k}{\sigma_k} \right)^2 (X_{\sigma})_{kj}. \end{aligned} \quad (32)$$

We can write the Hessian as a matrix-product as

$$H_P = X_{\sigma}^T \Lambda_1 X_{\mu} (X_{\mu}^T \Sigma^{-1} X_{\mu})^{-1} X_{\mu}^T \Lambda_1 X_{\sigma} + X_{\sigma}^T \Lambda_2 X_{\sigma} \quad (33)$$

with two $n \times n$ diagonal matrices

$$(\Lambda_1)_{ij} = \begin{cases} 0 & i \neq j \\ 2 \frac{y_i - \mu_i}{\sigma_i^2} & i = j \end{cases} \quad (\Lambda_2)_{ij} = \begin{cases} 0 & i \neq j \\ -2 \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 & i = j. \end{cases} \quad (34)$$

3 Distributions for estimators

Asymptotic theory of maximum-likelihood estimators tells that the vector of the combined estimators $(\hat{\beta}_{\mu}, \hat{\beta}_{\sigma})$ as defined in (8), is distributed approximately as

$$(\hat{\beta}_{\mu}, \hat{\beta}_{\sigma}) \sim \mathcal{N}_{k_{\mu} + k_{\sigma}} ((\beta_{\mu}^*, \beta_{\sigma}^*), \Sigma_{\beta\beta}) \quad \text{for } n \text{ large.} \quad (35)$$

This distribution is valid in the limit of a large number of observations n .

The covariance matrix $\Sigma_{\beta\beta}$ is given in terms of the inverse Fisher information matrix I_n :

$$\Sigma_{\beta\beta} = \frac{1}{n} I_n^{-1}. \quad (36)$$

The Fisher information matrix is given in terms of the expected value of the Hessian at $\beta_{\mu} = \beta_{\mu}^*$ and $\beta_{\sigma} = \beta_{\sigma}^*$:

$$I_n = -\frac{1}{n} E[H^*]. \quad (37)$$

The Hessian H is the Hessian of the full log-likelihood, in contrast to the profile-likelihood Hessian:

$$H^* = \begin{pmatrix} H_{\mu\mu}^* & H_{\mu\sigma}^* \\ H_{\mu\sigma}^* & H_{\sigma\sigma}^* \end{pmatrix} \quad (38)$$

with the three block-matrices defined as

$$(H_{\mu\mu}^*)_{ij} = \frac{\partial^2 \log L}{\partial \beta_{\mu i} \partial \beta_{\mu j}}, \quad (H_{\mu\sigma}^*)_{ij} = \frac{\partial^2 \log L}{\partial \beta_{\mu i} \partial \beta_{\sigma j}}, \quad (H_{\sigma\sigma}^*)_{ij} = \frac{\partial^2 \log L}{\partial \beta_{\sigma i} \partial \beta_{\sigma j}} \quad (39)$$

evaluated at $\beta_\mu = \beta_\mu^*$ and $\beta_\sigma = \beta_\sigma^*$.

We have already calculated $H_{\mu\mu}$ in (13). The other block matrices are given by

$$\begin{aligned} (H_{\mu\sigma}^*)_{ij} &= -2 \sum_{k=1}^n \frac{y_k - \mu_k^*}{\sigma_k^{*2}} (X_\mu)_{ki} (X_\sigma)_{kj} \\ (H_{\sigma\sigma}^*)_{ij} &= -2 \sum_{k=1}^n \left(\frac{y_k - \mu_k^*}{\sigma_k^*} \right)^2 (X_\sigma)_{ki} (X_\sigma)_{kj}. \end{aligned}$$

In matrix notation:

$$H_{\mu\mu}^* = -X_\mu^T \Sigma^{*-1} X_\mu, \quad H_{\mu\sigma}^* = -X_\mu^T \Lambda_1^* X_\sigma, \quad H_{\sigma\sigma}^* = X_\sigma^T \Lambda_2^* X_\sigma. \quad (40)$$

with Λ_1^* equal to Λ_1 with $\mu = \mu^*$ and $\sigma = \sigma^*$, and likewise for Λ_2^* .

When we take expected values and keep in mind that

$$\begin{aligned} E[Y - \mu^*] &= 0 \\ E[(Y_i - \mu_i^*)(Y_j - \mu_j^*)] &= \begin{cases} 0 & i \neq j \\ \sigma_i^{*2} & i = j \end{cases}, \end{aligned}$$

we arrive at

$$E[H_{\mu\mu}^*] = -X_\mu^T \Sigma^{*-1} X_\mu, \quad E[H_{\mu\sigma}^*] = 0, \quad E[H_{\sigma\sigma}^*] = -2X_\sigma^T X_\sigma \quad (41)$$

This brings the expected value of the Hessian in the form

$$E[H^*] = - \begin{pmatrix} X_\mu^T \Sigma^{*-1} X_\mu & 0 \\ 0 & 2X_\sigma^T X_\sigma \end{pmatrix}. \quad (42)$$

The function `fisher` in the `lmvar` package calculates the Fisher information matrix. It estimates $E[H^*]$ by replacing the true but unknown σ^* by its maximum-likelihood estimator $\hat{\sigma}$ in Σ^* .

The expectation value (42) brings the covariance matrix $\Sigma_{\beta\beta}$ in the form

$$\Sigma_{\beta\beta} = \begin{pmatrix} (X_\mu^T \Sigma^{*-1} X_\mu)^{-1} & 0 \\ 0 & \frac{1}{2} (X_\sigma^T X_\sigma)^{-1} \end{pmatrix}. \quad (43)$$

This implies that $\hat{\beta}_\mu$ and $\hat{\beta}_\sigma$ are independent stochastic variables distributed as

$$\begin{aligned}\hat{\beta}_\mu &\sim \mathcal{N}_{k_\mu}(\beta_\mu^*, (X_\mu^T \Sigma^{*-1} X_\mu)^{-1}) \\ \hat{\beta}_\sigma &\sim \mathcal{N}_{k_\sigma}(\beta_\sigma^*, \frac{1}{2} (X_\sigma^T X_\sigma)^{-1})\end{aligned}\quad \text{for } n \text{ large.} \quad (44)$$

We obtain for the asymptotic distribution of the maximum-likelihood estimators of μ^* and σ^*

$$\begin{aligned}\hat{\mu} &\sim \mathcal{N}_n(\mu^*, X_\mu (X_\mu^T \Sigma^{*-1} X_\mu)^{-1} X_\mu^T) \\ \log \hat{\sigma} &\sim \mathcal{N}_n(\log \sigma^*, \frac{1}{2} X_\sigma (X_\sigma^T X_\sigma)^{-1} X_\sigma^T)\end{aligned}\quad \text{for } n \text{ large.} \quad (45)$$

The expectation value and the variance for an element $\hat{\sigma}_i$ of $\hat{\sigma}$ are

$$\begin{aligned}E[\hat{\sigma}_i] &= \sigma_i^* \exp\left(\frac{(X_\sigma (X_\sigma^T X_\sigma)^{-1} X_\sigma^T)_{ii}}{4}\right) \\ \text{var}(\hat{\sigma}_i) &= (E[\hat{\sigma}_i])^2 \left(\exp\left(\frac{(X_\sigma (X_\sigma^T X_\sigma)^{-1} X_\sigma^T)_{ii}}{2}\right) - 1\right)\end{aligned}\quad \text{for } n \text{ large.} \quad (46)$$

The function `fitted.lmvar` (with the option `log = FALSE`) returns $\hat{\mu}$ and $\hat{\sigma}$.