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
$$Y \sim \mathcal{N}_n(\mu^\star, \Sigma) \tag{1}$$
in which $\mu^\star \in \mathbb{R}^n$ is the expected value and $\Sigma \in \mathbb{R}^{n,n}$ a diagonal covariance matrix
$$\Sigma_{ij} = \begin{cases} 0 & i \neq j \\ (\sigma_i^\star)^2 & i = j. \end{cases} \tag{2}$$
Assume that the vector of expectation values μ is linearly dependent on the values of the covariates in a model matrix X_μ:
$$\mu^\star = X_\mu \beta_\mu^\star \tag{3}$$
with $X_\mu \in \mathbb{R}^{n,k_\mu}$ and $\beta_\mu^\star \in \mathbb{R}^{k_\mu}$.

Similarly, assume that the vector $\sigma^\star = (\sigma_1^\star, \ldots, \sigma_n^\star)$ depends on the covariates in a model matrix X_σ as
$$\log \sigma^\star = X_\sigma \beta_\sigma^\star \tag{4}$$
where $\log \sigma^\star = (\log \sigma_1^\star, \ldots, \log \sigma_n^\star)$, $X_\sigma \in \mathbb{R}^{n,k_\sigma}$ and $\beta_\sigma^\star \in \mathbb{R}^{k_\sigma}$. 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 $\beta_\mu^\star$ and $\beta_\sigma^\star$, 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 $\beta_\mu^\star$ is defined uniquely only if X_μ is full-rank. If not, the space $\mathbb{R}^{k_\mu}$ can be split into subspaces such that there is a uniquely defined $\beta_\mu^\star$ 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 $\beta_\mu^\star$ 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 $\beta_\sigma^\star$ 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 $\beta_\mu^\star$ and $\beta_\sigma^\star$ 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

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

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

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

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

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

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:

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

These maximum likelihood estimators are taken to be the estimators of $\beta_\mu^\star$ and $\beta_\sigma^\star$. 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

$$\nabla_{\beta_\mu} \log \mathcal{L} = 0 \tag{9}$$

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

The derivatives are

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \beta_{\mu,i}} &= \sum_{k=1}^{n}(\frac{(y_k-\mu_k)}{\sigma_k^2})(X_\mu)_{ki} && (10)\\ &= \left(X_\mu^T \Lambda (y-\mu)\right)_i && (11) \end{aligned}$$

with Λ a diagonal matrix given by

$$\Lambda_{ij} = \frac{1}{\sigma_i^2}\delta_{ij}. \tag{12}$$

Hence

$$\nabla_{\beta_\mu} \log \mathcal{L} = X_\mu^T \Lambda (y-\mu) \tag{13}$$

and the maximum-likelihood equation (9) becomes

$$X_\mu^T \Lambda X_\mu \beta_\mu = X_\mu^T \Lambda y \tag{14}$$

which has the solution

$$\beta_\mu = \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \Lambda y. \tag{15}$$

Because of our assumption that X_μ is full rank, the inverse of the matrix $X_\mu^T \Lambda X_\mu$ can be taken.

It is easy to see that the solution (15) represents a maximum in the log-likelihood. The matrix $H_{\mu\mu}$ of second-order derivatives

$$(H_{\mu\mu})_{ij} = \frac{\partial^2 \log L}{\partial \beta_{\mu i} \partial \beta_{\mu j}} \tag{16}$$

is given by

$$H_{\mu\mu} = -X_\mu^T \Lambda X_\mu, \tag{17}$$

which is negative-definite for any β_σ.

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

$$\hat{\beta_\sigma} = \underset{\beta_\sigma \in \mathbb{R}^{k_\sigma}}{\operatorname{argmax}} \log \mathcal{L}_P(\beta_\sigma) \tag{18}$$

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

$$\mathcal{L}_P(\beta_\sigma) = \mathcal{L}(\beta_\mu(\beta_\sigma), \beta_\sigma). \tag{19}$$

with β_μ depending on β_σ as in (15).

To find $\hat{\beta}_\sigma$ from (18), we must solve

$$(\nabla_{\beta_\mu} \log \mathcal{L})\,(\nabla_{\beta_\sigma} \beta_\mu) + \nabla_{\beta_\sigma} \log \mathcal{L} = 0 \tag{20}$$

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

$$(\nabla_{\beta_\sigma} \beta_\mu)_{ij} = \frac{\partial \beta_{\mu i}}{\partial \beta_{\sigma j}}. \tag{21}$$

However, because of (9), the first term in (20) vanishes and we are left to solve

$$\nabla_{\beta_\sigma} \log \mathcal{L} = 0. \tag{22}$$

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

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial \beta_{\sigma i}} &= \sum_{k=1}^{n} (-(X_\sigma)_{ki} + \frac{(y_k - \mu_k)^2}{\sigma_k^2} (X_\sigma)_{ki}) \\ &= \sum_{k=1}^{n} (\frac{(y_k - \mu_k)^2}{\sigma_k^2} - 1)(X_\sigma)_{ki}. \end{aligned} \tag{23}$$

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

$$\nabla_{\beta_\sigma} \log \mathcal{L} = X_\sigma^T \lambda_\sigma \tag{24}$$

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

$$\lambda_{\sigma i} = \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 - 1. \tag{25}$$

The maximum-likelihood equations (22) take the form

$$X_\sigma^T \lambda_\sigma = 0. \tag{26}$$

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

$$\begin{aligned} \mu &= X_\mu \beta_\mu \\ &= X_\mu \left(X_\mu^T \Lambda X_\mu \right)^{-1} X_\mu^T \Lambda y \\ &= \Lambda^{1/2} X_\mu \left(X_\mu^T \Lambda X_\mu \right)^{-1} X_\mu^T \Lambda^{1/2} y \end{aligned} \tag{27}$$

where the latter form is the more symmetric, with

$$\left(\Lambda^{1/2} \right)_{ij} = \frac{1}{\sigma_i} \delta_{ij}. \tag{28}$$

The vector $(y - \mu)/\sigma$ can be written as

$$\frac{y - \mu}{\sigma} = \Lambda^{1/2} \left[I - \Lambda^{1/2} X_\mu \left(X_\mu^T \Lambda X_\mu \right)^{-1} X_\mu^T \Lambda^{1/2} \right] y \tag{29}$$

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 $\log \mathcal{L}_P$:

$$(H_P)_{ij} = \frac{\partial^2 \log \mathcal{L}_P}{\partial \beta_{\sigma j} \partial \beta_{\sigma i}} \tag{30}$$

Differentiation of (23) gives for the second-order derivatives

$$(H_P)_{ij} = -2 \sum_{k=1}^{n} (X_\sigma^T)_{ik} \frac{y_k - \mu_k}{\sigma_k^2} \left\{ \frac{\partial \mu_k}{\partial \beta_{\sigma j}} + (y_k - \mu_k)(X_\sigma)_{kj} \right\} \tag{31}$$

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 (15), the j-th column vector of the matrix is

$$
\begin{aligned}
\frac{\partial \mu}{\partial \beta_{\sigma j}} &= X_\mu \frac{\partial \beta_\mu}{\partial \beta_{\sigma j}} \\
&= X_\mu \left\{ \frac{\partial \left(X_\mu^T \Lambda X_\mu\right)^{-1}}{\partial \beta_{\sigma j}} X_\mu^T \Lambda + \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \frac{\partial \Lambda}{\partial \beta_{\sigma j}} \right\} y \\
&= X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} \left\{ -X_\mu^T \frac{\partial \Lambda}{\partial \beta_{\sigma j}} X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \Lambda + X_\mu^T \frac{\partial \Lambda}{\partial \beta_{\sigma j}} \right\} y \\
&= X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \frac{\partial \Lambda}{\partial \beta_{\sigma j}} \left\{ -X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \Lambda + I \right\} y \\
&= X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \frac{\partial \Lambda}{\partial \beta_{\sigma j}} (y - \mu) \qquad (32)
\end{aligned}
$$

The matrix $\partial \Lambda / (\partial \beta_{\sigma j})$ takes the form

$$
\begin{aligned}
\frac{\partial \Lambda}{\partial \beta_{\sigma j}} &= \sum_{i=1}^{n} \frac{\partial \Lambda}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \beta_{\sigma j}} \qquad (33) \\
&= -2 \begin{pmatrix} (X_\sigma)_{1j} & & 0 \\ & \ddots & \\ 0 & & (X_\sigma)_{nj} \end{pmatrix} \Lambda
\end{aligned}
$$

The j-th column vector of the matrix is

$$
\frac{\partial \mu}{\partial \beta_{\sigma j}} = -2 X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \begin{pmatrix} \frac{y_1 - \mu_1}{\sigma_1^2} (X_\sigma)_{1j} \\ \vdots \\ \frac{y_n - \mu_n}{\sigma_n^2} (X_\sigma)_{nj} \end{pmatrix} \qquad (34)
$$

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

$$
\frac{\partial \mu_k}{\partial \beta_{\sigma j}} = -2 \sum_{l=1}^{n} \left(X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \right)_{kl} \frac{y_l - \mu_l}{\sigma_l^2} (X_\sigma)_{lj}\,. \qquad (35)
$$

If we substitute this result in (31), we obtain for the element at row i and column j of the Hessian:

$$
\begin{aligned}
&(H_P)_{ij} = \\
&\quad 4 \sum_{k,l=1}^{n} (X_\sigma^T)_{ik} \frac{y_k - \mu_k}{\sigma_k^2} \left(X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \right)_{kl} \frac{y_l - \mu_l}{\sigma_l^2} (X_\sigma)_{lj} + \\
&\quad - 2 \sum_{k=1}^{n} (X_\sigma^T)_{ik} \left(\frac{y_k - \mu_k}{\sigma_k} \right)^2 (X_\sigma)_{kj}. \qquad (36)
\end{aligned}
$$

We can write the Hessian as a matrix-product as

$$
H_P = X_\sigma^T \tilde{\Lambda}_1 X_\mu \left(X_\mu^T \Lambda X_\mu\right)^{-1} X_\mu^T \tilde{\Lambda}_1 X_\sigma + X_\sigma^T \tilde{\Lambda}_2 X_\sigma \qquad (37)
$$

with two $n \times n$ diagonal matrices

$$\begin{aligned} \left(\tilde{\Lambda}_1\right)_{ij} &= 2\,\frac{y_i - \mu_i}{\sigma_i^2}\,\delta_{ij} \\ \left(\tilde{\Lambda}_2\right)_{ij} &= -2\left(\frac{y_i - \mu_i}{\sigma_i}\right)^2\,\delta_{ij}. \end{aligned} \tag{38}$$

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

$$(\hat{\beta}_\mu, \hat{\beta}_\sigma) \sim \mathcal{N}_{k_\mu + k_\sigma}\left((\beta_\mu^\star, \beta_\sigma^\star), \Sigma_{\beta\beta}\right) \qquad \text{for } n \text{ large.} \tag{39}$$

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:

$$\Sigma_{\beta\beta} = \frac{1}{n} I_n^{-1}. \tag{40}$$

The Fisher information matrix is given in terms of the expected value of the Hessian:

$$I_n = -\frac{1}{n} E[H]. \tag{41}$$

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

$$H = \begin{pmatrix} H_{\mu\mu} & H_{\mu\sigma} \\ H_{\mu\sigma}^T & H_{\sigma\sigma} \end{pmatrix} \tag{42}$$

with the three block-matrices defined as

$$(H_{\mu\mu})_{ij} = \frac{\partial^2 \log L}{\partial\beta_{\mu i}\partial\beta_{\mu j}},\ (H_{\mu\sigma})_{ij} = \frac{\partial^2 \log L}{\partial\beta_{\mu i}\partial\beta_{\sigma j}},\ (H_{\sigma\sigma})_{ij} = \frac{\partial^2 \log L}{\partial\beta_{\sigma i}\partial\beta_{\sigma j}}. \tag{43}$$

We have already calculated $H_{\mu\mu}$ in (17). Differentiation of (10) and (23) respectively gives

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

In matrix notation:

$$H_{\mu\sigma} = X_\mu^T \Lambda_1 X_\sigma, \qquad H_{\sigma\sigma} = X_\sigma^T \Lambda_2 X_\sigma, \tag{44}$$

with the diagonal matrices

$$(\Lambda_1)_{ij} = -2\,\frac{y_i - \mu_i}{\sigma_i^2}\,\delta_{ij}, \qquad (\Lambda_2)_{ij} = -2\left(\frac{y_i - \mu_i}{\sigma_i}\right)^2 \delta_{ij}. \tag{45}$$

When we take expected values, we have to take $\beta_\mu = \beta_\mu^\star$ and $\beta_\sigma = \beta_\sigma^\star$. Keeping in mind that

$$E[Y - \mu^\star] = 0$$
$$E[(Y_i - \mu_i^\star)(Y_j - \mu_j^\star)] = \sigma_i^{\star 2}\delta_{ij},$$

we have

$$E[H_{\mu\mu}] = -X_\mu^T \Lambda^\star X_\mu,\ E[H_{\mu\sigma}] = 0,\ E[H_{\sigma\sigma}] = -2X_\sigma^T X_\sigma \tag{46}$$

where $\Lambda^\star$ is Λ with σ taken to be $\sigma^\star$. This brings the expected value of the Hessian in the form

$$E[H] = -\begin{pmatrix} X_\mu^T \Lambda^\star X_\mu & 0 \\ 0 & 2X_\sigma^T X_\sigma \end{pmatrix}. \tag{47}$$

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

The expectation value (47) brings the covariance matrix $\Sigma_{\beta\beta}$ in the form

$$\Sigma_{\beta\beta} = \begin{pmatrix} \left(X_\mu^T \Lambda^\star X_\mu\right)^{-1} & 0 \\ 0 & \frac{1}{2}\left(X_\sigma^T X_\sigma\right)^{-1} \end{pmatrix}. \tag{48}$$

This implies

$$\begin{aligned} \hat{\beta}_\mu &\sim \mathcal{N}_{k_\mu}(\beta_\mu^\star,\ \left(X_\mu^T \Lambda^\star X_\mu\right)^{-1}) \\ \hat{\beta}_\sigma &\sim \mathcal{N}_{k_\sigma}(\beta_\sigma^\star,\ \tfrac{1}{2}\left(X_\sigma^T X_\sigma\right)^{-1}) \end{aligned} \qquad \text{for } n \text{ large.} \tag{49}$$

We obtain for the asymptotic distribution of the maximum-likelihood estimators of $\mu^\star$ and $\sigma^\star$

$$\begin{aligned} \hat{\mu} &\sim \mathcal{N}_n(\mu^\star,\ X_\mu \left(X_\mu^T \Lambda^\star X_\mu\right)^{-1} X_\mu^T) \\ \log\hat{\sigma} &\sim \mathcal{N}_n(\log\sigma^\star,\ \tfrac{1}{2} X_\sigma \left(X_\sigma^T X_\sigma\right)^{-1} X_\sigma^T) \end{aligned} \qquad \text{for } n \text{ large.} \tag{50}$$

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

$$\begin{aligned} E[\hat{\sigma}_i] &= \sigma_i^\star \exp\left(\frac{(X_\sigma \left(X_\sigma^T X_\sigma\right)^{-1} X_\sigma^T)_{ii}}{4}\right) \\ \mathrm{var}(\hat{\sigma}_i) &= (E[\hat{\sigma}_i])^2 \left(\exp(\frac{(X_\sigma \left(X_\sigma^T X_\sigma\right)^{-1} X_\sigma^T)_{ii}}{2}) - 1\right) \end{aligned} \qquad \text{for } n \text{ large.} \tag{51}$$

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