Errata (1st edition)

- Ch. 3, p. 97: The regression coefficient for $x_{i,c-1}$ should not have an index i:
$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{c-1} x_{i,c-1} + \ldots + \varepsilon_i.$$
- Ch. 4, p. 180: The weighted least squares criterion if falsely abbreviated as GLS:
$$\mathrm{WLS}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X\beta}) = \sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2.$$
- Ch. 4, p. 180: Missing inverse in the derivation of the covariance matrix of $\boldsymbol{\varepsilon}^*$:
$$\mathrm{Cov}(\boldsymbol{\varepsilon}^*) = \mathrm{E}(\boldsymbol{W}^{1/2}\boldsymbol{\varepsilon\varepsilon}'\boldsymbol{W}^{1/2}) = \sigma^2 \boldsymbol{W}^{1/2}\boldsymbol{W}^{-1}\boldsymbol{W}^{1/2} = \sigma^2 \boldsymbol{I}.$$
- Ch. 4, p. 187: Wrong model specification in Example 4.2:
Extending model (4.3), we therefore assume the variance model
$$\sigma_i^2 = \sigma^2\, h(\alpha_0 + \alpha_1\, areao_i + \alpha_4 yearco_i + \alpha_5 yearco2_i + \alpha_6 yearco3_i),$$
where again *yearco*, *yearco2*, and *yearco3* are cubic orthogonal polynomials for year of construction (see Example 3.5 on p. 90). Based on this model, we obtain $T = 1164.37$ as the Breusch-Pagan test statistic.
- Ch. 4, p. 187: Mistake in the weights for two-stage least squares:
$$\hat{w}_i = \boldsymbol{z}_i'\hat{\boldsymbol{\alpha}}.$$
- Ch. 4, p. 188: Mistake in the weights for two-stage least squares:
$$\hat{w}_i = \exp(\boldsymbol{z}_i'\hat{\boldsymbol{\alpha}}).$$
- Ch. 4, p. 188: Mistake in the weights for two-stage least squares:
$$\hat{w}_i = \exp(\hat{\eta}_i)$$
- Ch. 4, p. 188: Mistake in the covariance matrix for two-stage least squares:
$$\widehat{\mathrm{Cov}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\boldsymbol{X}'\mathrm{diag}\left(\frac{1}{\hat{w}_1}, \ldots, \frac{1}{\hat{w}_n}\right)\boldsymbol{X})^{-1},$$
- Ch. 4, p. 209: Mistakes in the"'estimation equations": Omitting their derivation they are given by
$$2\boldsymbol{X}'\boldsymbol{X\beta} - 2\boldsymbol{X}'\boldsymbol{y} + \lambda\,\mathrm{sign}(\boldsymbol{\beta}) = \boldsymbol{0}$$
where the sign function is applied elementwise to the vector of regression coefficients.
- Ch. 4, p. 218: The lack of fit has to be assessed with the *negative* derivative:
When starting with initial guesses $\hat{\boldsymbol{\beta}}^{(0)}$, we can compute the lack of fit information associated with this starting values as the corresponding negative derivative of the least squares criterion, i.e.,
$$-\frac{\partial}{\partial\boldsymbol{\beta}}\,\mathrm{LS}(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}} = -2\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(0)}\right).$$

- Ch. 4, p. 219: The formula
$$\hat{b}_j = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)u_i}{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}$$
does not work for the intercept ($j = 0$) where the denominator would be equal to zero.
- Ch. 4, p. 223: The lack of fit has to be assessed with the *negative* derivative:
If a candidate predictor value $\hat{\boldsymbol{\eta}}^{(0)}$ is given, the corresponding lack of fit can then be evaluated with $\mathrm{LS}(\hat{\boldsymbol{\eta}}^{(0)})$, but more detailed information is contained in the unit-specific negative gradients
$$u_i = -\frac{\partial}{\partial \eta_i}\,\mathrm{LS}(\boldsymbol{\eta})\bigg|_{\eta_i = \hat{\eta}_i^{(0)}} = 2\left(y_i - \hat{\eta}_i^{(0)}\right).$$
For a perfect fit, all these gradients will be zero while large negative gradients point towards observations where the fit could be substantially improved. In fact, the gradients are basically the residuals obtained by plugging in the candidate predictor (multiplied with a factor of 2).
- Ch. 4, p. 227: Errors in the discussion of the inverse gamma prior for σ^2:
Of particular interest is the case $a = b$ and both values approaching zero. Then the distribution converges to an improper distribution that also results from a general prior construction principle (Jeffreys' prior), see for example Held & Sabanés Bové (2012). Another interesting case is when $a = 1$ and b is chosen small. In this case, the distribution of $\log(\sigma^2)$ tends to a uniform distribution as can be shown analytically through the change in variables theorem ...
- Ch. 4, p. 233: Missing 0.5 for the parameter a of the NIG prior:
In case of a noninformative prior with $\boldsymbol{m} = \mathbf{0}$, $\boldsymbol{M}^{-1} = \mathbf{0}$, $a = -p/2$, and $b = 0$, we obtain...
- Ch. 4, p. 242: Missing indices in the hyperparameters:
  - For the ridge prior draw $(\tau^2)^{(t)} \,|\, \cdot$ from an inverse gamma distribution with parameters
  $$a^{new} = a_{\tau^2} + \frac{k}{2}, \qquad b^{new} = b_{\tau^2} + \frac{1}{2}\tilde{(}\boldsymbol{\beta}^{(t)})'\tilde{\boldsymbol{\beta}}^{(t)}.$$
  - For the LASSO prior, sample
  $$\begin{aligned}(1/\tau_j^2)^{(t)} \,|\, \cdot &\sim \mathrm{InvGauss}(\tfrac{|\lambda^{(t-1)}|}{|\beta_j^{(t)}|}, (\lambda^{(t-1)})^2),\\ (\lambda^2)^{(t)} \,|\, \cdot \quad &\sim \mathrm{G}\left(a_\lambda + k, b_\lambda + \tfrac{1}{2}\sum_{j=1}^{k}(\tau_j^2)^{(t)}\right).\end{aligned}$$
- Ch. 4, p. 253: At various places in Section 4.4.4, σ^2 should be replaced by τ^2:
$$\beta_j \,|\, \delta_j, \sigma^2 \sim (1 - \delta_j)\mathrm{Dirac}(0) + \delta_j\ \mathrm{N}(0, \sigma^2). \tag{4.36}$$
$$\beta_j \,|\, \delta_j, \tau^2 \sim (1 - \delta_j)\,\mathrm{N}(0, \nu_0\,\tau^2) + \delta_j\ \mathrm{N}(0, \tau^2), \tag{4.37}$$
The same changes have to be made in Box 4.10.

- Ch. 4, p. 253: Missing "=0" in the probability statement:
  ... i.e., $\mathrm{P}(\delta_j = 1) = \theta$ and $\mathrm{P}(\delta_j = 0) = 1 - \theta$.
- Ch. 4, p. 259: The argument of $\mathrm{LS}(\cdot)$ should be $\boldsymbol{\beta}$ instead of $\hat{\boldsymbol{\beta}}$:

$$\mathrm{LS}(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$$

- Ch. 4, p. 262: Missing index $\boldsymbol{\beta}$ for the covariance matrix:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left(\frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{K}\right)^{-1}.$$

- Ch. 4, p. 264: Missing closing bracket:

$$p(\omega_j \mid \cdot) \propto \left(\frac{1}{\omega_j^3}\right)^{1/2} \exp\left(-\frac{\lambda^2}{2\mu\,\omega_j}(\omega_j - \mu)^2\right),$$

- Ch. 4, p. 267: Missing index τ^2:
  For $\delta_j = 0$, we have to exchange b_{τ^2} by $\nu_0 b_{\tau^2}$ and arrive at ...
- Ch. 4, p. 267: Equality sign should be distributed as:

$$\tau_j^2 \mid \cdot \sim (1 - \delta_j)\mathrm{IG}(a_{\tau^2} + 1/2, \nu_0 b_{\tau^2} + 1/2\beta_j^2) + \delta_j\,\mathrm{IG}(a_{\tau^2} + 1/2, b_{\tau^2} + 1/2\beta_j^2).$$

- Ch. 4, p. 275: Numerator and denominator have to be switched to obtain the formula for the multiplicative interpretation on odds ratios:

$$\frac{\mathrm{P}(y_i = 1 \mid x_{i1} + 1, \dots)}{\mathrm{P}(y_i = 0 \mid x_{i1} + 1, \dots)} \Big/ \frac{\mathrm{P}(y_i = 1 \mid x_{i1}, \dots)}{\mathrm{P}(y_i = 0 \mid x_{i1}, \dots)} = \exp(\beta_1).$$

- Ch. 5, p. 283: Index i runs from 1 to G:

$$\bar{y}_i \sim \mathrm{B}(n_i, \pi_i)/n_i, \qquad i = 1, \dots, G$$

- Ch. 5, p. 285: Misplaced transpose in the formula for the Fisher information:

$$\boldsymbol{F}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'/\sigma^2 = \frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X}.$$

- Ch. 5, p. 288: Smaller p-values indicate a lack of fit:
  Larger values in the observed test statistic indicate lack of fit and therefore correspond to smaller p-values.
- Ch. 5, p. 292: Missing bar in the variance formula

$$\mathrm{Var}(\bar{y}_i) = \phi\frac{\pi_i(1 - \pi_i)}{n_i}.$$

- Ch. 5, p. 292: Mistakes in the estimates for the dispersion parameter:

$$\hat{\phi}_P = \frac{1}{G - p}\chi^2 \quad \text{or} \quad \hat{\phi}_D = \frac{1}{G - p}D,$$

  where G is the number of groups.

- Ch. 5, p. 2)5: Mistake in the normalising constant of the likelihood:
  …apart from the additive constant $-\sum_i \log(y_i!)$ …
- Ch. 5, p. 297: Mistake in the Pearson statistic:

$$\chi^2 = \sum_{i=1}^{G} \frac{(\bar{y}_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i / n_i}$$

- Ch. 5, p. 297: Mistakes in the estimates for the dispersion parameter:

$$\hat{\phi}_P = \frac{1}{G-p}\chi^2 \quad \text{or} \quad \hat{\phi}_D = \frac{1}{G-p} D\,.$$

- Ch. 5, p. 301: Missing sign in the reciprocal link:

$$\mu_i = -\frac{1}{\eta_i} = -\frac{1}{\boldsymbol{x}_i'\boldsymbol{\beta}}.$$

- Ch. 5, p. 308: Mistakes in the estimate for the dispersion parameter:
  Using the variance function, the dispersion parameter can then be estimated consistently by

$$\hat{\phi} = \frac{1}{G-p} \sum_{i=1}^{G} \frac{(\bar{y}_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)/w_i},$$

  where p denotes the number of regression parameters, $\hat{\mu}_i = h(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$ is the estimated expectation, $v(\hat{\mu}_i)$ is the estimated variance function, the weights w_i are given by n_i, and the data should be grouped as much as possible.
- Ch. 5, p. 308: Mistakes in the definition of the Pearson statistic:

$$\chi^2 = \sum_{i=1}^{G} \frac{(\bar{y}_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)/w_i}$$

- Ch. 5, p. 309: lq instead of lr for the likelihood ratio
  We reject H_0 when

$$lr, w, u > \chi^2_r(1-\alpha).$$

- Ch. 5, p. 314: Duplicated negative sign in the formula for the Fisher information matrix:

$$\boldsymbol{F}_p(\boldsymbol{\beta}) = \mathrm{E}\left(-\frac{\partial^2 \log(p(\boldsymbol{\beta}\,|\,\boldsymbol{y}))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) = \boldsymbol{F}(\boldsymbol{\beta}) + \boldsymbol{M}^{-1}\,.$$

- Ch. 6, p. 327: Mistake in the probability function for the multinomial distribution based on m independent trials:

$$f(\boldsymbol{y}\,|\,\boldsymbol{\pi}) = \frac{m!}{y_1! \cdot \ldots \cdot y_c!(m - y_1 - \ldots - y_c)!}\pi_1^{y_1} \cdot \ldots \cdot \pi_c^{y_c}(1 - \pi_1 - \ldots - \pi_c)^{m-y_1-\ldots-y_c}.$$

- Ch. 6, p. 346: Mistake in the formula for the score function:

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{X}'\boldsymbol{D}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \quad \boldsymbol{F}(\boldsymbol{\beta}) = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$$

  where $\boldsymbol{\mu} = (\ldots, n_i\boldsymbol{\pi}_i', \ldots)'$.

- Ch. 7, p. 374: Mistakes in the elements of $\boldsymbol{V}^{-1}$:
  More specifically, the elements on the main diagonal of $\boldsymbol{V}_i^{-1}$ are given by
  $$\frac{\sigma^2 + (n_i - 1)\tau_0^2}{\sigma^2(\sigma^2 + n_i\tau_0^2)},$$
  and the elements above and below the main diagonal are
  $$-\frac{\tau_0^2}{\sigma^2(\sigma^2 + n_i\tau_0^2)}.$$
- Ch. 7, p. 385: Missing condition in the distribution of the random effects:
  $$p(\boldsymbol{y} \,|\, \boldsymbol{\vartheta}) \propto \int p(\boldsymbol{y} \,|\, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\vartheta})\, p(\boldsymbol{\gamma}|\boldsymbol{\vartheta})\, p(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, d\boldsymbol{\gamma} \tag{7.37}$$
- Ch. 8, p. 427: Mistake in the index set:
  $$B_j^0(z) = I(\kappa_j \le z < \kappa_{j+1}) = \begin{cases} 1 & \kappa_j \le z < \kappa_{j+1} \\ 0 & \text{otherwise} \end{cases} \quad j = 1, \ldots, d,$$
- Ch. 8, p. 458: Mistake in the explanation of the nonparametric smoothing interpretation of kriging:
  We thus obtain the representation
  $$\boldsymbol{y} = \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}$$
  with $\tilde{\boldsymbol{Z}}[i, j] = \rho(|z_i - z_{(j)}|)$ and $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_d)'$.
- Ch. 8, p. 460: Inconsistent definition of neighborhoods:
  In general, observations are considered to be their own neighbors, so that $i \in N(z_i)$ holds.
- Ch. 8, p. 470: Missing "hats" in Box. 8.11:
  3. Simultaneous confidence bands at $\{z_1, \ldots, z_j, \ldots, z_r\}$ based on the joint distribution of $(\hat{f}(z_1), \ldots, \hat{f}(z_r))'$:
     $$\hat{f}(z_j) \pm m_{1-\alpha}\sigma\sqrt{\boldsymbol{s}(z_j)'\boldsymbol{s}(z_j)},$$
     where the quantile $m_{1-\alpha}$ of Eq. (8.18) is determined via simulation.
- Ch. 8, p. 517: Mistake in the transposed signs in the joint normal distribution:
  $$\begin{pmatrix} \boldsymbol{y} \\ \eta_0 \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta} \\ \mu_0 \end{bmatrix}, \begin{bmatrix} \tau^2 \boldsymbol{Z}\boldsymbol{R}\boldsymbol{Z}' + \sigma^2 \boldsymbol{I}_n & \tau^2 \boldsymbol{r} \\ \tau^2 \boldsymbol{r}' & \tau^2 \end{bmatrix}\right)$$
- Ch. 9, p. 570: Mistake in the formula for the covariance matrix:
  $$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \mathrm{Cov}(\boldsymbol{\beta} \,|\, \cdot) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1},$$
- Ch. 10, p. 603: Mistake in the definition of $(x)_+$:
  … where $(x)_+ = \max(x, 0)$ and …

- Ch. 10, p. 603: Mistake in the matrix representation of the model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_\tau + \boldsymbol{u}_\tau - \boldsymbol{v}_\tau,$$

- Ch. 10, p. 610: Sloppy notation for the flat prior of $\boldsymbol{\beta}_\tau$:
  We thereby assume noninformative priors $p(\boldsymbol{\beta}_\tau) \propto$ const and the usual inverse gamma priori for σ^2, i.e., $\sigma^2 \sim \mathrm{IG}(a, b)$ with hyperparameters a and b.
- Ch. 10, p. 611: Sloppy notation for the flat prior of $\boldsymbol{\beta}_\tau$ in Box 10.2:

$$p(\boldsymbol{\beta}_\tau) \propto \text{const}$$
$$z_i \mid \sigma^2 \sim \mathrm{Expo}(1/\sigma^2)$$
$$\sigma^2 \sim \mathrm{IG}(a, b).$$

- Ch. 10, p. 618: Mistake in the rewritten optimality criterion:

$$\begin{aligned} \mathrm{E}\left(w_\tau(y)|y-q|\right) &= \int_{-\infty}^{\infty} w_\tau(y)|y-q|f(y)dy \\ &= \int_{-\infty}^{q} (1-\tau)(y-q)f(y)dy - \int_{q}^{\infty} \tau(y-q)f(y)dy. \end{aligned}$$

- Ch. 10, p. 618: Incorrect formatting of the subscript for z_1:

$$\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{\beta}_\tau, \sigma^2 \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}_\tau + \xi\boldsymbol{z},\ \sigma^2\boldsymbol{W}^{-1}), \qquad \boldsymbol{W} = \mathrm{diag}(w_1, \dots, w_n), \qquad \boldsymbol{z} = (z_1, \dots, z_n)'$$

- Ch. 10, p. 619: Missing part in the full conditional for σ^2:

$$a' = a + \frac{3n}{2} \text{ and } b' = b + \frac{1}{2}(\boldsymbol{X}\boldsymbol{\beta}_\tau + \xi\boldsymbol{z} - \boldsymbol{y})'\boldsymbol{W}(\boldsymbol{X}\boldsymbol{\beta}_\tau + \xi\boldsymbol{z} - \boldsymbol{y}) + \sum_{i=1}^{n} z_i.$$

- App. A, p. 626: Mistake in the definition of row and column space:
  The column space $C(\boldsymbol{A})$ of an $n \times p$-matrix is the subspace of $\mathbb{R}^n$ spanned by the columns of $\boldsymbol{A}$, i.e.

$$C(\boldsymbol{A}) := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} \text{ for some } \boldsymbol{y} \in \mathbb{R}^p\}.$$

  The row space $R(\boldsymbol{A})$ is defined correspondingly.
- App. B, p. 641: Right and left truncation are defined the wrong way:
  For $a = -\infty$ or $b = \infty$, X is said to be right or left truncated, respectively.