UNIVERSITY OF WARWICK

Uses of {gnm} for Generalized (Non)linear Modelling

Heather Turner

November 26, 2025

https://hturner.github.io/BIBC

Background

{gnm} (Turner and Firth, 2007) was released on CRAN 20 years ago 🎉

Time to

- Review the original motivation for {gnm}
- Explore ways {gnm} has been used in practice

Photo by Shaylyn on Unsplash

https://hturner.github.io/BIBC

Introduction to GNMs and {gnm}

Generalized Nonlinear Models (GNMs)

In a Generalized *Linear* Model (GLM) we have

$$g(E(y_i)) = g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$$

and

$$\text{Var}(y_i) = \phi a_i V(\mu_i)$$

A GNM extends this to

$$g(\mu_i) = \eta(\boldsymbol{x}_i; \boldsymbol{\beta})$$

where $\eta(\boldsymbol{x}_i; \boldsymbol{\beta})$ is nonlinear in the parameters $\boldsymbol{\beta}$.

Motivation

GNMs may be thought of as…

… an extension of Nonlinear Least Squares

- using a link function to model a non-Gaussian response

… an extension of GLMs

- using nonlinear functions of parameters to produce a more parsimonious model and interpretable model.

{gnm} focuses on the latter

Classic Models that are GNMs

- Goodman's RC model for 2 way tables (Goodman, 1979, *JASA*)

$$\log \mu_{rc} = \alpha_r + \beta_c + \gamma_r \delta_c$$

- Stereotype model for ordered categorical data (Anderson, 1984, *JRSSB*)

$$pr(y_i = c|\boldsymbol{x}_i) = \frac{\exp(\beta_{0c} + \gamma_c \boldsymbol{\beta}^T \boldsymbol{x}_i)}{\sum_r \exp(\beta_{0r} + \gamma_r \boldsymbol{\beta}^T \boldsymbol{x}_i)}$$

- Lee-Carter age-specific mortality model (Lee and Carter, 1992, *JASA*)

$$\log(\mu_{ay}/e_{ay}) = \alpha_a + \beta_a \gamma_y$$

Poisson models to analyse case-crossover studies

Effect of Ozone Pollution on Deaths in London

Armstrong et al (2014, *BMC Medical Research Methodology*) provide the following example data original from Bhaskaran et al, 2013, *International Journal of Epidemiology*:

- `date` daily from 2002-01-01 to 2006-12-31
- `ozone` daily average, $\mu g/m^3$
- `temperature` $^\circ C$
- `numdeaths`

```
1 library(foreign)
2 london <- read.dta("londondataset2002_2006.dta")
```

Case cross-over data

Define month-weekday strata:

record	date	weekday	stratum	numdeaths	ozone	temperature
1	2002-01-01	Tue	2002-Jan-Tue	199	4.59	-0.23

Expand each record to make case-control sets within the strata:

record	date	weekday	case	weight	ozone	temperature
1	2002-01-01	Tue	1	199	4.59	-0.23
1	2002-01-08	Tue	0	199	3.13	3.51
1	2002-01-15	Tue	0	199	15.95	7.24
1	2002-01-22	Tue	0	199	43.09	9.04
1	2002-01-29	Tue	0	199	38.11	10.91

Conditional logistic model

Given that one case occurs in each stratum, we can model the event $D_{i,s}$ that the death in stratum s occurs on day i as follows:

$$D_{i,s} \sim \text{Bernoulli}\left(\pi_i = \frac{\exp\left\{\boldsymbol{\beta}^T \boldsymbol{x}_i\right\}}{\sum_{j \in s(i)} \exp\left\{\boldsymbol{\beta}^T \boldsymbol{x}_j\right\}}\right)$$

In our London example, we have:

- 8034 rows: 1826 dates × 4-5 of each weekday per month
- 2 parameters: one for `ozone` and one for `temperature`

Equivalent Poisson model

Armstrong et al (2014) propose to fit the equivalent Poisson model to $y_{i,s}$, the number of deaths on date i in stratum s, where

$$E(y_i) = \exp(\alpha_s + \boldsymbol{\beta}^T \boldsymbol{x}_i)$$

and α_s are nuisance parameters that ensure the multinomial denominators are matched.

In this case we have

- 1826 rows, one for each date
- 2 parameters of interest plus $5 \times 12 \times 7 = 420$ nuisance parameters!

Fitting GLMs

GLMs are estimated with an IWLS algorithm, where

$$\boldsymbol{\beta}^{(r+1)} = \left(\boldsymbol{X}^T \boldsymbol{W}^{(r)} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}^{(r)} \boldsymbol{z}^{(r)}$$

For GNMs, {gnm}:

- Uses a **generalized inverse** to avoid specifying identifiability constraints for nonlinear terms.
- Can **eliminate** parameters of a nuisance factor from $\boldsymbol{X}$, for efficient estimation.

Augmented Least Squares

For an ordinary least squares model

$$\left[(\boldsymbol{y}|\boldsymbol{X})^T(\boldsymbol{y}|\boldsymbol{X})\right]^{-1} = \begin{pmatrix} \boldsymbol{y}^T\boldsymbol{y} & \boldsymbol{y}^T\boldsymbol{X} \\ \boldsymbol{X}^T\boldsymbol{y} & \boldsymbol{X}^T\boldsymbol{X} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$$

where $\boldsymbol{A}_{11}$, $\boldsymbol{A}_{12}$ and $\boldsymbol{A}_{22}$ are functions of $\boldsymbol{y}^T\boldsymbol{y}$, $\boldsymbol{X}^T\boldsymbol{y}$ and $\boldsymbol{X}^T\boldsymbol{X}$.

It can be shown that

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = -\frac{\boldsymbol{A}_{21}}{\boldsymbol{A}_{11}}$$

requiring only the first row (column) of the inverse to be found.

Application to Nuisance Parameters I

Let

$$\tilde{\boldsymbol{X}} = \boldsymbol{W}^{\frac{1}{2}}(\boldsymbol{z}|\boldsymbol{X}) = (\boldsymbol{U}|\boldsymbol{V})$$

- $\boldsymbol{V}$ is the $nk \times n$ matrix of dummy variables corresponding to the nuisance factor with n levels.
- $\boldsymbol{U}$ is a $nk \times (p+1)$ matrix where p is the number of model parameters.

Typically $\boldsymbol{n} >> \boldsymbol{p}$

Application to Nuisance Parameters II

Then

$$(\tilde{\boldsymbol{X}^T \boldsymbol{X}})^- = \begin{pmatrix} \boldsymbol{U}^T\boldsymbol{U} & \boldsymbol{U}^T\boldsymbol{V} \\ \boldsymbol{V}^T\boldsymbol{U} & \boldsymbol{V}^T\boldsymbol{V} \end{pmatrix}^- = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix}$$

Again, only the first row (column) of this generalised inverse is required to estimate $\hat{\boldsymbol{\beta}}$, so we are only interested in $\boldsymbol{B}_{11}$ and $\boldsymbol{B}_{12}$.

$$\begin{aligned} \boldsymbol{B}_{11} &= (\boldsymbol{U}^T\boldsymbol{U} - \boldsymbol{U}^T\boldsymbol{V}(\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T\boldsymbol{U})^- \\ \boldsymbol{B}_{12} &= -(\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T\boldsymbol{U}\boldsymbol{B}_{11} \end{aligned}$$

Elimination of the Nuisance Factor

The structure of $\tilde{\boldsymbol{X}}$ makes things simpler

- $\boldsymbol{U}^T\boldsymbol{U}$ is $(p+1)\times(p+1)$: not expensive to compute.
- $\boldsymbol{V}^T\boldsymbol{V}$ is $n \times n$ diagonal: the non-zero elements can be computed directly.
- $\boldsymbol{V}^T\boldsymbol{U}$ is $n \times (p+1)$: equivalent to aggregating the rows of $\boldsymbol{U}$ by levels of the nuisance factor.

So we do not need to construct the large $nk \times n$ matrix $\boldsymbol{V}$.

Application to the Case-Crossover Poisson Model

Fit with standard GLM function

```
tic()
poisson_glm <- glm(numdeaths ~ ozone + temperature + stratum,
                   data = london, family = poisson)
toc()
```

```
1.198 sec elapsed
```

```
tic()
poisson_gnm <- gnm(numdeaths ~ ozone + temperature, eliminate = stratum,
                   data = london, family = poisson)
toc()
```

```
0.007 sec elapsed
```

```
coef(poisson_gnm)
```

```
Coefficients of interest:
       ozone  temperature
0.0003384861 0.0041931648
```

Explore further

Methodology and London example:

- Armstrong et al (2014) Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis, BMC Medical Research Methodology [~250 citations]

Some recent Australasian applications:

- Chen et al (2022) Ambient air pollution and epileptic seizures: A panel study in Australia, *Epilepsia*
- Campbell et al (2024) Assessing mortality associated with heatwaves in the cool climate region of Tasmania, Australia, *The Journal of Climate Change and Health*

Models for longitudinal categorical data

https://hturner.github.io/BIBC

Longitudinal categorical data

We have repeated measures of a categorical outcome for multiple subjects, e.g.

Subject	Time 1	Time 2	Time 3
1	None	Mild	Moderate
2	None	None	Severe
3	Mild	Moderate	Moderate

Let $y_{it} \in \{1, \ldots, J\}$ be the response for subject i at time t.

We want to model the marginal probabilities

$$P(y_{it} = j | \boldsymbol{x}_{it})$$

regarding the correlation between repeated measures as a nuisance.

Indicator vector representation

Represent the categorical response for subject i at time t as a vector of integers:

$$\boldsymbol{y}_{it} = (y_{i1}, \ldots, y_{iJ})^\top$$

where

$$y_{ij} = \begin{cases} 1, & \text{if } y_i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Then the expected value is the vector of category probabilities we want to model

$$E[\boldsymbol{y}_{it}|\boldsymbol{x}_{it}] = (P(y_{it} = 1|\boldsymbol{x}_{it}),\ P(y_{it} = 2|\boldsymbol{x}_{it}),\ \ldots,\ P(y_{it} = J|\boldsymbol{x}_{it}))^\top = \boldsymbol{\pi}_{it}$$

Generalized Estimating Equation Approach

1. Model the marginal mean $E[\boldsymbol{y}_{it} | \boldsymbol{x}_{it}] = \boldsymbol{\pi}_{it}(\boldsymbol{\beta})$
2. Stack the repeated outcomes for a subject: $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, \ldots, \boldsymbol{y}_{iT}^\top)^\top$
3. Specify a "working" covariance matrix: $\mathrm{Var}(\boldsymbol{y}_i) = \boldsymbol{V}_i(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$
4. Solve the estimating equation

$$\boldsymbol{U}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) = \sum_i \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\pi}_i) = \mathbf{0}, \quad \mathbf{D}_i = \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}^\top}$$

- **No need to specify the full joint distribution**
- **Gives consistent estimates of β even if α mis-specified**

Marginal Model

A univariate multinomial model can provide the required model for $\boldsymbol{\pi}_{it}$, e.g.

- Baseline-category multinomial logit model for nominal outcomes

$$\log\left(\frac{\pi_{itj}}{\pi_{itJ}}\right) = \beta_{j0} + \boldsymbol{\beta}_j^\top \mathbf{x}_{it}, \qquad j = 1, \ldots, J-1$$

- Cumulative link model for ordinal outcomes

$$F^{-1}\left[P(y_{it} \leq j \mid \mathbf{x}_{it})\right] = \beta_{j0} + \boldsymbol{\beta}^\top \mathbf{x}_{it}, \qquad j = 1, \ldots, J-1$$

from which π_{itj} can be derived as differences between cumulative probabilities.

Modelling the Covariance Matrix V_i

Any model for V_i will result in consistent estimates for $\boldsymbol{\beta}$, however, if the model is

- Too simple: lose efficiency
- Too complex: convergence problems, imprecise estimation of $\boldsymbol{\alpha}$

Want a parsimonious representation that captures the correlation patterns

Touloumis et al (2013, *Biometrics*) propose to use **association models** for the marginalized contingency tables, implemented in R package {multgee}.

Touloumis et al Approach

1. For each time pair $g = (t, t')$ create a 2-way contingency table of frequencies $f_{jj'g}$.
2. Model the 3-way contingency table with the homogeneous RC-G(1) model:

$$\log f_{jj'g} = \lambda + \lambda_j^R + \lambda_{j'}^C + \lambda_g^G + \lambda_{jg}^{RG} + \lambda_{j'g}^{CG} + \phi^g \mu_j^g \mu_{j'}^g$$

3. Define $\boldsymbol{\alpha}$ as the vector of all local odds ratios, $\theta_{jj'g}$, based on

$$\begin{bmatrix} f_{jj'g} & f_{j(j'+1)g} \\ f_{(j+1)j'g} & f_{(j+1)(j'+1)g} \end{bmatrix}$$

   i.e. $\log(\theta_{jj'g}) = \phi^g(\mu_j^g - \mu_{j+1}^g)(\mu_{j'}^g - \mu_{j'+1}^g)$

4. Estimate $P(y_{it} = j, y_{it'} = j' | x_i)$ by iterative proportional fitting to calculate V_i

Alternative Association Models

Different models for $\log(\theta_{jj'g})$ can be obtained through different models for $\phi^g \mu_j^g \mu_{j'}^g$:

Structure	Model for Log Local Odds Ratio	Interaction Formula	Response
Independent	1	N/A	Any
Uniform	ϕ	`~ r:c`	Ordinal
Category exchangeable	ϕ^g	`~ G:r:c`	Ordinal
Time exchangeable	$\phi(\mu_j - \mu_{j+1})(\mu_{j'} - \mu_{j'+1})$	`~ MultHomog(R,C)`	Any
Time-specific RC	$\phi^g(\mu_j^g - \mu_{j+1}^g)(\mu_{j'}^g - \mu_{j'+1}^g)$	`~ MultHomog(G:R, G:C)`	Any

Where `MultHomog()` is a function in {gnm} for specifying homogeneous multiplicative interactions ($\mu_r \mu_c$).

Example: Housing data

The `housing` data from {multgee} is an example of nominal longitudinal data:

- `y`: housing in streets or shelters (0), community housing (1), or independent housing (2)
- `time`: 0, 6, 12, and 24 months
- `sec`: binary indicator of access to Section 8 Rental Certificate (financial support)

```
1 library(multgee)
2 data(housing)
3 head(housing, 4)
```

```
  id y time sec
1  1 1    0   1
2  1 2    6   1
3  1 2   12   1
4  1 2   24   1
```

Modelling with Local Odds Ratios GEE

Estimate time-specific intrinsic parameters ϕ^g (in a RC-G(1) model):

```
phi <- intrinsic.pars(y, data = housing, id = id, repeated = time, rscale = "nominal")
phi
```

```
[1] 1.0644358 0.8748243 0.6171560 2.5821515 1.6521935 3.6752211
```

There is a wide range in the strength of association, so should use time-specific RC model

```
fit_exch <- nomLORgee(y ~ factor(time) * sec, data = housing, id = id, repeated = time)
fit_rc <- update(fit_exch, LORstr = "RC")
fit_ind <- update(fit_exch, LORstr = "independence")
gee_criteria(fit_ind, fit_exch, fit_rc)[,1:3]
```

```
              QIC      CIC     RJC
fit_ind  1533.624 1533.624 16.0000
fit_exch 1517.530 1518.330 15.6001
fit_rc   1509.793 1510.677 15.5579
```

Exploring further

The housing data is analysed in:

- Touloumis, A. (2011) GEE for multinomial responses. PhD dissertation, University of Florida.

The model is described in:

- Touloumis, A. et al (2013) GEE for multinomial responses using a local odds ratios parameterization *Biometrics* [~150 citations]

A recent Australasian example:

- Cooper, IL et al (2025) Returning to work following parental leave: the experiences of Australian anaesthetists, Anaesthesia and Intensive Care.

Modelling the mean-variance relationship for in-vitro diagnostic assays

https://hturner.github.io/BIBC

Performance of In-Vitro Diagnostic (IVD) Assays

Manufacturers of IVD assays need to provide information the precision of their product.

This is based on a precision experiment. `CA19_9` from {VCA} is an example from the Clinical and Laboratory Standards Institute guidelines.

6 samples were assayed with 5 replicates on 5 days, at 3 sites

- `result`: measured concentration of CA19-9 (tumour marker)
- `day`: day 1, 2, 3, 4, or 5
- `site`: laboratory 1, 2 or 3
- `sample`: P1, P2, P5, Q3, Q4, Q6 (“P” patient sample pool, “Q” control, number indicates concentration)

Variance Component Analysis

`anovaVCA()` from {VCA} computes variance components based on ANOVA:

```
1 library(VCA)
2 data(CA19_9)
3 anovaVCA(result ~ site/day, Data = subset(CA19_9, sample == "P1"))
```

```
Result Variance Component Analysis:
-----------------------------------

  Name      DF         SS         MS         VC        %Total    SD        CV[%]
1 total     11.318142                       1.086864  100       1.042528  8.629244
2 site      2          22.041867  11.020933  0.384291  35.357751 0.619912  5.131154
3 site:day  12         16.964     1.413667   0.177773  16.356539 0.421632  3.489944
4 error     60         31.488     0.5248     0.5248    48.28571  0.724431  5.996282

Mean: 12.08133 (N = 75)

Experimental Design: balanced  |  Method: ANOVA
```

Precision Performance Table

Doing this for all samples, we can obtain a “Precision Performance Table”

	Mean	Reproducibility %CV	Between-Site %CV	Between-Day %CV	Repeatability %CV
sample.P1	12.08	8.6	5.1	3.5	6.0
sample.P2	41.58	4.4	3.1	0.8	3.1
sample.Q3	55.75	4.1	3.2	1.3	2.2
sample.Q4	165.70	3.8	3.3	0.8	1.7
sample.P5	379.10	2.4	1.3	0.5	2.0
sample.Q6	414.30	3.7	3.1	0.4	2.1

where *Reproducibility* is the total variation and *Repeatability* is the pure assay imprecision.

Precision Profiles

Precision profiles model the relationship between the (component of) variance and the mean response.

{VFP} implements several models as in the Variance Function Program software:

Number	Model	Type
1	$\sigma^2 = 1$	linear
6	$\sigma^2 = \beta_1 + \beta_2\mu + \beta_3\mu^J$	nonlinear
7	$\sigma^2 = \beta_1 + \beta_2\mu^J$	nonlinear
8	$\sigma^2 = (\beta_1 + \beta_2\mu)^J$	nonlinear
9	$\sigma^2 = \beta_1\mu^J$	nonlinear

plus special cases Models 2-5 with J set to 2 or another specified value.

GNMs for Precision Profiles

For balanced designs we have

$$E(\hat{\sigma}_c^2) = \sigma_c^2 \qquad \text{Var}(\hat{\sigma}_c^2) \approx \frac{2}{\nu_c}\sigma_c^4$$

where ν_c is the degrees of freedom for $\hat{\sigma}_c^2$.

Therefore we can fit precision profiles using a Gamma GNM with an identity link

$$E(\hat{\sigma}_c^2) = \eta(\mu, \boldsymbol{\beta}) \qquad \text{Var}(\hat{\sigma}_c^2) = \frac{\phi}{\alpha_c}E(\hat{\sigma}_c^2)^2$$

with weights $\alpha_c = \nu_c/2$.

Modelling Total Variability

First use `anovaVCA()` to estimate variance components for each sample:

```
vca <- anovaVCA(result ~ site/day, Data = CA19_9, by = "sample")
```

Then use `get_mat()` to get the mean, a total variance (`VC`), corresponding degrees of freedom (`DF`) and the coefficient of variation (as %, `CV`) for all samples:

```
library(VFP)
total <- get_mat(vca, vc = "total")
total
```

	Mean	DF	VC	SD	CV
sample.P1	12.08133	11.318142	1.086864	1.042528	8.629244
sample.P2	41.58400	7.604586	3.376848	1.837620	4.419056
sample.Q3	55.74667	4.896189	5.257296	2.292879	4.113034
sample.Q4	165.65600	3.331477	39.752635	6.304969	3.806061
sample.P5	379.09067	16.709246	85.059893	9.222792	2.432872
sample.Q6	414.28667	4.112871	241.089499	15.527057	3.747902

Example: Model 7

```
powfun7 <- function (x) {
    list(predictors = list(beta1 = 1, beta2 = 1, J = 1),
         variables = list(substitute(x)),
         term = function(predictors, variables) {
            paste0(predictors[1], "+",
                   predictors[2], "*", variables[1], "^", predictors[3])
         })
}
class(powfun7) <- "nonlin"
mod7 <- gnm(VC ~ powfun7(Mean) - 1, family = Gamma(link = "identity"), weights = DF/2,
                data = total, start = c(1, 1, 2), verbose = FALSE)
coef(mod7)
```

```
Coefficients:
     beta1      beta2          J
0.71739105 0.00563237 1.66887107
```

```
deviance(mod7)
```

```
[1] 1.758392
```

Selecting Variance Function

`fit_vfp()` will fit all 9 models and select best fitting by AIC

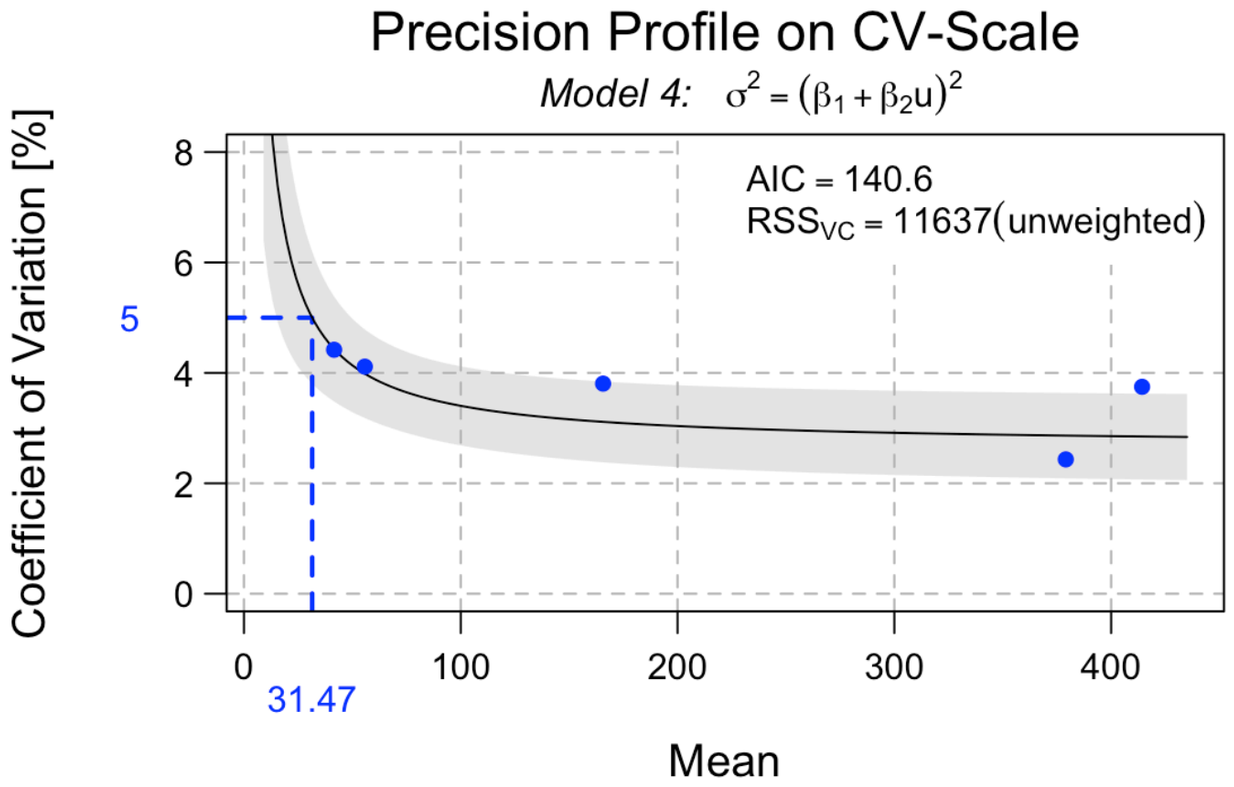
```
1 all_mod <- fit_vfp(total, quiet = TRUE)
2 all_mod
```

```
(VFP) Variance-Function
-----------------------

Model 4*:   sigma^2=(beta1+beta2*u)^2

Coefficients:
  beta1   beta2
 0.7329 0.02671

AIC = 140.6  RSS = 11637  Deviance = 1.875 GoF P-value= 1.000

*Best-fitting model of 8 VFP-models overall
```

Functional sensitivity

Determine concentrations at which a specified CV is not exceeded.

```
plot(all_mod, model.no = 4, type = "cv", ylim = c(0, 8), Prediction = 5)
```



https://hturner.github.io/BIBC

Explore Further

The VFP vignette gives further detail on the use of precision profiles.

{VFP} aims to implement the functionality of the Variance Function Program distributed by the Australasian Association for Clinical Biochemistry and Laboratory Medicine

- The VFP documentation by W. A. Sadler (formerly Christchurch hospital) gives several references relating to the methods and application.

Summary

{gnm} provides a unified framework for a wide range of models

Applications “in the wild” have used several of its features:

- The “eliminate” feature for efficient handling of nuisance parameters
- In-built “nonlin” functions to specify multiplicative terms
- Custom “nonlin” functions for specialized models

For more on {gnm} see the vignette or the BIBC 2025 tutorial

https://hturner.github.io/BIBC

References

Turner, H and D Firth (2007) gnm: A Package for Generalized Nonlinear Models *R News*

Anderson, J. A. (1984). Regression and Ordered Categorical Variables *JRSS B*

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories *JASA*

Lee, R. D. and L. Carter (1992). Modelling and forecasting the time series of {US} mortality *JASA*