# Models

Kasper Schou Telkamp

Section for Dynamical Systems

**DTU Compute**
Department of Applied Mathematics and Computer Science

# Outline

- Data exploration
  - VTEC / STEC
- Hierachical Poisson-Normal model
  - Formulation
  - Implementation
  - Results
- Hierachical Poisson-Gamma model
  - Formulation
  - Probability function for $Y$
  - Inference on individual group means
  - Implementation
  - Results

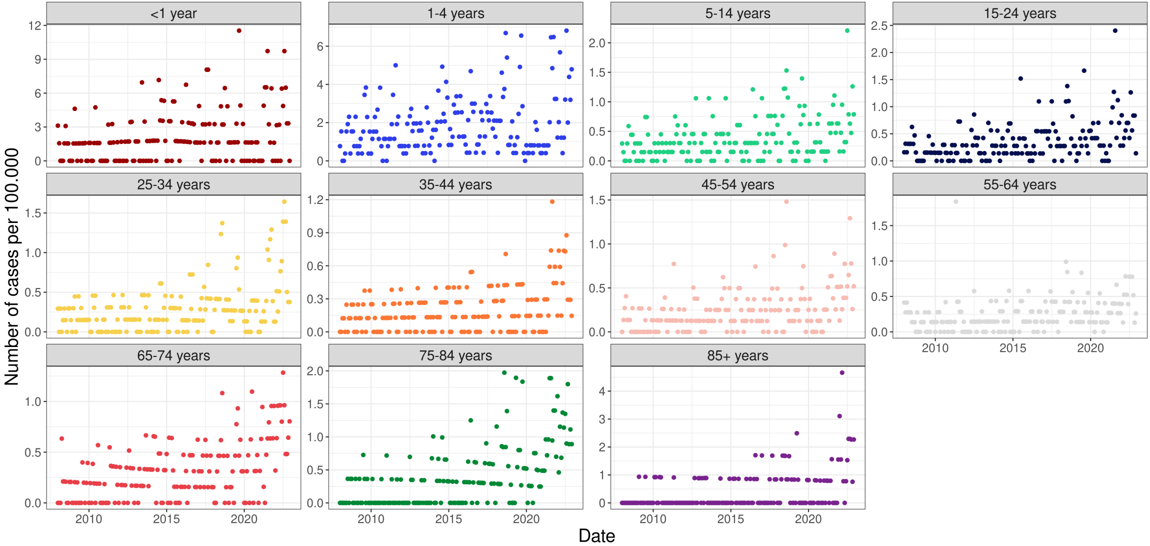
Data exploration

# VTEC / STEC

| Date | ageGroup | $y_{it}$ | $x_{it}$ |
|---|---|---|---|
| 2008-01-01 | $<$1 year | 2 | 64137 |
| 2008-01-01 | 1-4 years | 2 | 259910 |
| 2008-01-01 | 5-14 years | 2 | 680529 |
| 2008-01-01 | 15-24 years | 1 | 631724 |
| ... | ... | ... | ... |
| 2022-12-01 | 55-64 years | 2 | 773073 |
| 2022-12-01 | 65-74 years | 5 | 621965 |
| 2022-12-01 | 75-84 years | 4 | 449423 |
| 2022-12-01 | 85+ years | 3 | 132181 |

# VTEC / STEC

Shiga- og veratoxin producerende E. coli.



# Formulation

The count observations are assumed to follow a Poisson distribution with intensities $\lambda_{it}$. Also, we shall assume that the count is proportional to the population size, $x_{it}$, within each age group, $i$, at a given time point, $t$. Hence, in terms of the canonical link for the Poisson distribution the model is

$$\log(\lambda_{it}) = \mathbf{X}_i^T \beta_{it} + \log(x_{it}) + u_{it} \tag{1}$$

Here $\mathbf{X}_i$ is a vector of ones, and $\beta_{it}$ contains the corresponding fixed effect parameter. The random effects $u_{it}$ are assumed to be Gaussian.

$$u_{it} = \epsilon_{it} \tag{2}$$

where $\epsilon_{it} \sim \mathrm{N}(0, \sigma^2)$ is a white noise process, and $\sigma$ is a model parameter.

# Fomulation

Henceforth, the model can be formulated as a two-level hierarchical model

$$Y_{it}|u_{it} \sim \mathrm{Pois}(\lambda_{it}) \tag{3a}$$

$$u_{it} \sim \mathrm{N}(0, \sigma^2) \tag{3b}$$

# Implementation - Objective function in C++

```cpp
#include <TMB.hpp>              // Links in the TMB libraries

template<class Type>
Type objective_function<Type>::operator() ()
{
  DATA_VECTOR(y);                                 // Data vector transmitted from R
  DATA_VECTOR(w)                          // Data vector transmitted from R
  DATA_FACTOR(ageGroup);                  // Data factor transmitted from R

  PARAMETER_VECTOR(u);                        // Random effects

  // Parameters
  PARAMETER_VECTOR(log_lambda);           // Parameter value transmitted from R
  PARAMETER(log_sigma_u);                 // Parameter value transmitted from R

  vector<Type> lambda  = exp(log_lambda);
  Type sigma_u = exp(log_sigma_u);

  int nobs = y.size();
  Type mean_ran = Type(0);

  int i;

  Type f = 0;                             // Declare the "objective function"
  for(int t=0; t < nobs; t++){
    i = ageGroup[t];
    f -= dnorm(u[t],mean_ran,sigma_u,true);
    f -= dpois(y[t],exp(lambda[i]-log(w[t]))*exp(u[t]),true);
  }

  return f;
}
```

# Implementation - Call from R

```r
# Import libraries
library(readr)
library(dplyr)
library(TMB)

# Import the data
dat <- read_rds(file = "../../data/processed/dat.rds")

# Only consider some of the data
y <- dat %>%
  filter(caseDef == "Shiga- og veratoxin producerende E. coli.") %>%
  group_by(Date, ageGroup) %>%
  reframe(y = sum(cases), n = sum(n))

compile(file = "PoissonLognormal.cpp")  # Compile the C++ file
# dyn.unload(dynlib("PoissonLognormal"))
dyn.load(dynlib("PoissonLognormal"))    # Dynamically link the C++ code

# Function and derivative
PoisLN <- MakeADFun(
  data = list(y = y$y, ageGroup = y$ageGroup, w = y$n),
  parameters = list(u = rep(1, length(y$y)),
                    log_lambda = rep(log(1), nlevels(y$ageGroup)),
                    log_sigma_u = log(1)),
  random = "u",
  DLL = "PoissonLognormal"
)

opt <- nlminb(start = PoisLN$par, PoisLN$fn, PoisLN$gr, lower = c(0.01, 0.01))
```

# Results

| Parameter | Estimate | Std. Error |
|---|---|---|
| $\lambda_{<1year}$ | 2.38 | 0.01 |
| $\lambda_{1-4years}$ | 2.61 | 0.01 |
| $\lambda_{5-14years}$ | 2.63 | 0.01 |
| $\lambda_{15-24years}$ | 2.64 | 0.01 |
| $\lambda_{25-34years}$ | 2.62 | 0.01 |
| $\lambda_{35-44years}$ | 2.61 | 0.01 |
| $\lambda_{45-54years}$ | 2.63 | 0.01 |
| $\lambda_{55-64years}$ | 2.62 | 0.01 |
| $\lambda_{65-74years}$ | 2.61 | 0.01 |
| $\lambda_{75-84years}$ | 2.53 | 0.01 |
| $\lambda_{85+years}$ | 2.35 | 0.01 |
| $\sigma$ | 1.01 | 0.03 |

# Results


Shiga- og veratoxin producerende E. coli.
Hierarchical Poisson Normal model
<1 year
1-4 years
5-14 years
15-24 years
25-34 years
35-44 years
45-54 years
55-64 years
65-74 years
75-84 years
85+ years
Random effect, $u_{it}$
Date
2010
2015
2020
2
1
0
-1

# Formulation

$$Y_{it} | u_{it} \sim \text{Pois}(\lambda_i u_{it}) \tag{4a}$$

$$u_{it} \sim \text{G}(1/\beta_i, \beta_i) \tag{4b}$$

# Probability function for $Y$

$$
\begin{aligned}
P[Y=y] &= g_Y(y;\lambda,\beta) \\
&= \frac{\lambda^y}{y!\Gamma(1/\beta)\beta^{1/\beta}} \frac{\beta^{y+1/\beta}\Gamma(y+1/\beta)}{(\lambda\beta+1)^{y+1/\beta}} \\
&= \frac{\Gamma(y+1/\beta)}{\Gamma(1/\beta)y!} \frac{1}{(\lambda\beta+1)^{1/\beta}} \left(\frac{\lambda\beta}{\lambda\beta+1}\right)^y \\
&= \binom{y+1/\beta-1}{y} \frac{1}{(\lambda\beta+1)^{1/\beta}} \left(\frac{\lambda\beta}{\lambda\beta+1}\right)^y \ , \ for\ y=0,1,2,\dots
\end{aligned}
\tag{5}
$$

where we have used the convention

$$
\binom{z}{y} = \frac{\Gamma(z+1)}{\Gamma(z+1-y)y!} \tag{6}
$$

The marginal distribution of $Y$ is a negative binomial distribution, $Y \sim \mathrm{NB}\left(1/\beta, 1/(\lambda\beta+1)\right)$

# Proof

The probability function for the conditional distribution of $Y$ for given $u$

$$f_{Y|u}(y;\lambda,u) = \frac{(\lambda u)^y}{y!}\exp(-\lambda u) \tag{7}$$

and the probability density function for the distribution of $u$ is

$$f_u(u;\beta) = \frac{1}{\beta\Gamma(1/\beta)}\left(\frac{u}{\beta}\right)^{1/\beta-1}\exp(-u/\beta) \tag{8}$$

# Proof

Given (7) and (8), the probability function for the marginal distribution of $Y$ is determined from

$$
\begin{aligned}
g_Y(y;\lambda,\beta) &= \int_{u=0}^{\infty} f_{Y|u}(y;\lambda,u) f_u(u;\beta)\,du \\
&= \int_{u=0}^{\infty} \frac{(\lambda u)^y}{y!}\exp(-\lambda u)\frac{1}{\beta\Gamma(1/\beta)}\left(\frac{u}{\beta}\right)^{1/\beta-1}\exp(-u/\beta)\,du \\
&= \frac{\lambda^y}{y!\Gamma(1/\beta)\beta^{1/\beta}}\int_{u=0}^{\infty} u^{y+1/\beta}\exp\big(-u(\lambda\beta+1)/\beta\big)\,du
\end{aligned}
\tag{9}
$$

# Proof

In (9) it is noted that the integrand is the *kernel* in the probability density function for a Gamma distribution, $\mathrm{G}\left(y+1/\beta, \beta/(\lambda\beta+1)\right)$. As the integral of the density shall equal one, we find by adjusting the norming constant that

$$\int_{u=0}^{\infty} u^{y+1/\beta} \exp\left(-u\Big/\Big(\beta/(\lambda\beta+1)\Big)\right) du = \frac{\beta^{y+1/\beta}\Gamma(y+1/\beta)}{(\lambda\beta+1)^{y+1/\beta}} \tag{10}$$

and then (5) follows

# Inference on individual group means

Consider the hierarchical Poisson-Gamma model in (4), and assume that a value $Y = y$ has been observed. Then the conditional distribution of $u$ for given $Y = y$ is a Gamma distribution,

$$u|Y = y \sim \mathrm{G}\left(y + 1/\beta, \beta/(\lambda\beta + 1)\right) \tag{11}$$

with mean

$$\mathrm{E}[u|Y = y] = \frac{y\beta + 1}{\lambda\beta + 1} \tag{12}$$

and variance

$$\mathrm{V}[u|Y = y] = \frac{(y\beta^2 + \beta)}{(\lambda\beta + 1)^2} \tag{13}$$

# Proof

The conditional distribution is found using Bayes Theorem

$$
\begin{aligned}
g_u(u|Y=y) &= \frac{f_{y,u}(y,u)}{g_Y(y;\lambda,\phi)} \\
&= \frac{f_{y|u}(y;u)g_u(u)}{g_Y(y;\lambda,\beta)} \\
&= \frac{1}{g_Y(y;\lambda,\beta)}\left(\frac{(\lambda u)^y}{y!}\exp(-\lambda u)\frac{1}{\beta\Gamma(1/\beta)}\left(\frac{u}{\beta}\right)^{1/\beta-1}\exp(-u/\beta)\right) \\
&\propto u^{y+1/\beta-1}\exp\big(-u(\lambda\beta+1)/\beta\big)
\end{aligned}
\tag{14}
$$

We identify the *kernel* of the probability density function

$$
u^{y+1/\beta-1}\exp(-u(\lambda\beta+1)/\beta) \tag{15}
$$

as the kernel of a Gamma distribution, $\mathrm{G}(y+1/\beta,\beta/(\lambda\beta+1))$

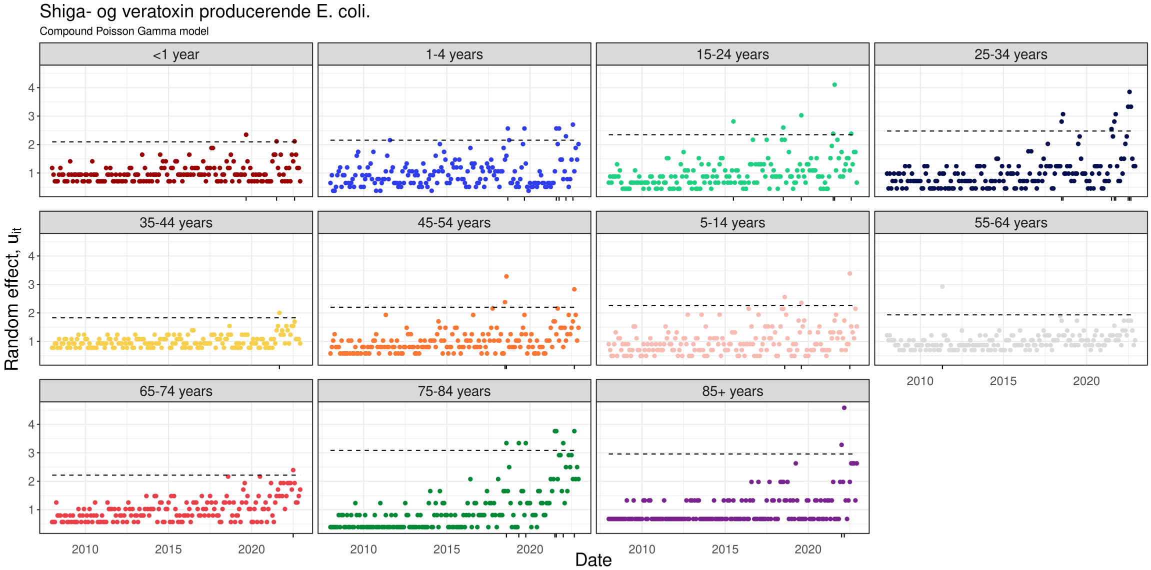
# Implementation - Define negative likelihood function in R

```r
# Define the negative likelihood function for the marginal distribution of Y
nll.age <- function(theta, data){
  # Extract counts
  y <- data$y
  # Extract agegroups
  ageGroup <- data$ageGroup
  # Extract number of agegroups
  n.ageGroup <- n_distinct(data$ageGroup)

  # Define parameters
  lambda <- theta[1:n.ageGroup]
  beta <- theta[(n.ageGroup+1):(n.ageGroup*2)]

  # Construct the size and probability for the negative binomial distribution
  r <- 1/beta
  p <- 1/(lambda*beta+1)

  # Initilize the log-likelihood
  ll <- 0
  for(i in 1:nrow(data)){
    ll = ll + dnbinom(x = y[i],
                      size = r[ageGroup[i]],
                      prob = p[ageGroup[i]],
                      log = TRUE)
  }

  # Return the negative log-likelihood
  -ll
}
```

# Results

| ageGroup | $\lambda$ | $\beta$ |
|---|---|---|
| <1 year | 1.24 | 0.33 |
| 1-4 years | 4.57 | 0.36 |
| 5-14 years | 2.47 | 0.42 |
| 15-24 years | 2.57 | 0.48 |
| 25-34 years | 2.08 | 0.57 |
| 35-44 years | 1.47 | 0.20 |
| 45-54 years | 1.88 | 0.39 |
| 55-64 years | 1.77 | 0.25 |
| 65-74 years | 1.89 | 0.40 |
| 75-84 years | 1.45 | 1.08 |
| 85+ years | 0.50 | 0.97 |

# Results


Shiga- og veratoxin producerende E. coli.
Compound Poisson Gamma model
<1 year
1-4 years
15-24 years
25-34 years
35-44 years
45-54 years
5-14 years
55-64 years
65-74 years
75-84 years
85+ years
Random effect, $u_{it}$
Date
2010
2015
2020
1
2
3
4