Semiparametric Vector Generalized Linear Models.

Estimation and Computation

Gabriel Dennis

June 7, 2023

Outline of talk

Introduction

Motivating Example

Proposed Model

Fitting the Model

Applications

Table of Contents

GLMs

A Generalized Linear Model (GLM) for modelling data

$$(Y_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^q, i = 1, \ldots, n$$

consist of three components

GLMs

A Generalized Linear Model (GLM) for modelling data

$$(Y_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^q, i = 1, \ldots, n$$

consist of three components

- A random component: a reference distributions for Y which originates from an exponential family.

GLMs

A Generalized Linear Model (GLM) for modelling data

$$(Y_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^q, i = 1, \dots, n$$

consist of three components

- A random component: a reference distributions for Y which originates from an exponential family.
- A systematic component: a linear predictor $\eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$

GLMs

A Generalized Linear Model (GLM) for modelling data

$$(Y_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^q, i = 1, \ldots, n$$

consist of three components

- A random component: a reference distributions for Y which originates from an exponential family.
- A systematic component: a linear predictor $\eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$
- A link function: A function which maps $\mathbb{E}[Y_i \mid \boldsymbol{X}_i] = \mu_i$ to η_i. In most cases it is more convenient to use the inverse link function $g^{-1}(\mu_i) = \eta_i$.

Common GLMS

Common GLMS include

Common GLMS

Common GLMS include

- Normal Linear Model: Which assumes $Y_i \sim \mathrm{N}(\mu_\mathrm{i}, \sigma^2)$ and $g(\mu_i) = \eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$.

Common GLMS

Common GLMS include

- Normal Linear Model: Which assumes $Y_i \sim \mathrm{N}(\mu_\mathrm{i}, \sigma^2)$ and $g(\mu_i) = \eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$.
- Poisson Regression: Which assumes $Y_i \sim \mathrm{Po}(\mu_\mathrm{i})$ and $g(\mu_i) = \log(\boldsymbol{X}_i^T \boldsymbol{\beta})$.

Common GLMS

Common GLMS include

- Normal Linear Model: Which assumes $Y_i \sim \mathrm{N}(\mu_{\mathrm{i}}, \sigma^2)$ and $g(\mu_i) = \eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$.
- Poisson Regression: Which assumes $Y_i \sim \mathrm{Po}(\mu_{\mathrm{i}})$ and $g(\mu_i) = \log(\boldsymbol{X}_i^T \boldsymbol{\beta})$.
- Logistic Regression: Which assumes $Y_i \sim$ Ber(μ_i) and $g(\mu_i) = \frac{\log(\boldsymbol{X}_i^T \boldsymbol{\beta})}{1-\log(\boldsymbol{X}_i^T \boldsymbol{\beta})}$.

VGLMs

A VGLM (Vector Generalized Linear Model) aims to generalize GLMs to multivariate responses, however, multivariate generalizations of common families of probability distributions (e.g., poisson, gamma, ...) are difficult to construct.

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship
- observed data often exhibit over-or under-dispersion relative to the postulated model

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship
- observed data often exhibit over-or under-dispersion relative to the postulated model

For multivariate data, problems arise when specifying a parametric distribution and the response covariance structure. For most cases there is no standard parametric model which could produce a given dataset.

Table of Contents

Motivating Example (Butterfly Dataset)

This dataset contains 66 counts of 14 species of butterflies in Boulder Colorado, USA.

Table 1: Snippet of Butterfly Counts for the 3 most common species (Hui et al. 2013; Oliver, Prudic and Collinge 2006)

site	Pieris.rapae	Colias.philodice	Colias.eurytheme	building	vegetation	habitat
1	0	0	0	2.1	10.8	mixed
2	1	1	2	2.1	10.8	mixed
3	1	1	2	19.8	1.7	mixed
4	0	2	1	5.3	0	mixed
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

We would like to fit a joint model, as species could complement or compete with each other within a site.

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

We would like to fit a joint model, as species could complement or compete with each other within a site.

However, no multivariate generalisation of the poisson distribution allows for both positive and negatively correlated response components.

Table of Contents

Vector Generalisation

This model is a vector generalisation of a Semiparametric Generalized Linear Model [1], which is based on a exponential tilt reformulation of the joint distribution from which the data originates for which the response distribution F remains unspecified.

[1](P. J. Rathouz and Gao 2009; Huang and P. Rathouz 2012; Huang 2014)

Vector Generalisation

This model is a vector generalisation of a Semiparametric Generalized Linear Model [1], which is based on a exponential tilt reformulation of the joint distribution from which the data originates for which the response distribution F remains unspecified.

This allows arbitrary nonparametric response distribution parameter F and the mean model parameters $\boldsymbol{\beta}$ to be simultaneously estimated without loss of information.

[1](P. J. Rathouz and Gao 2009; Huang and P. Rathouz 2012; Huang 2014)

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\}\ dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\}\ dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Here

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\}\; dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Here

- b_i are each observations normalising constant,

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\}\ dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Here

- b_i are each observations normalising constant,
- $dF_i(\boldsymbol{y})$ is an exponential tilt of the response distribution $dF(\boldsymbol{y})$,

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\}\ dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Here

- b_i are each observations normalising constant,
- $dF_i(\boldsymbol{y})$ is an exponential tilt of the response distribution $dF(\boldsymbol{y})$,
- where the amount of tilting $\boldsymbol{\theta}$ determined by the mean $\boldsymbol{\mu}(\boldsymbol{X}_i^T \boldsymbol{\beta})$ of the observation $\boldsymbol{Y}_i$.

Model Tilt Parameters

The normalising constants $b_i = b(\boldsymbol{X}_i, \boldsymbol{\beta}, F)$ are defined to be

$$b(\boldsymbol{X}_i, \boldsymbol{\beta}, F) = -\log\left\{\int_{\mathcal{Y}} \exp\{\boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y})\right\},$$

Model Tilt Parameters

The normalising constants $b_i = b(\boldsymbol{X}_i, \boldsymbol{\beta}, F)$ are defined to be

$$b(\boldsymbol{X}_i, \boldsymbol{\beta}, F) = -\log\left\{\int_{\mathcal{Y}} \exp\{\boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y})\right\},$$

and the vector of tilt parameters $\boldsymbol{\theta}_i = \boldsymbol{\theta}(\boldsymbol{X}, \boldsymbol{\beta}, F) \in \mathbb{R}^K$ is the implicit solution to

$$\mu_{(k)}(\boldsymbol{X}_{(k)}^T \boldsymbol{\beta}_{(k)}) = \int_{\mathcal{Y}} y_{(k)} \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y}), \quad k = 1, \ldots, K$$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F|\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F | \boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

One way to estimate the densities $dF(\boldsymbol{Y}_i)$ is by replacing them with the histogram estimators p_i, which are assigned to values in the observed support $\{\boldsymbol{Y}_i \in \mathbb{R}^k | i = 1, 2, \ldots, n\}$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F|\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

One way to estimate the densities $dF(\boldsymbol{Y}_i)$ is by replacing them with the histogram estimators p_i, which are assigned to values in the observed support $\{\boldsymbol{Y}_i \in \mathbb{R}^k | i = 1, 2, \ldots, n\}$

This produces empirical log-likelihood.

$$\ell(\boldsymbol{\beta}, \boldsymbol{p}) = \sum_{i=1}^{n} \log p_i + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i$$

Estimation of β and $\boldsymbol{p}$

$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{p}})$ are then the joint maximisers of the empirical log likelihood.

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{p}}) = \arg\max \ell(\boldsymbol{\beta}, \boldsymbol{p})$$

Estimation of β and $\boldsymbol{p}$

Subject to empirical analogous of the normalising constraints

$$1 = \sum_{i=1}^{n} p_i \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\}, \quad j = 1, 2, 3, \ldots, n$$

Estimation of β and p

Subject to empirical analogous of the normalising constraints

$$1 = \sum_{i=1}^{n} p_i \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\}, \quad j = 1, 2, 3, \dots, n$$

and the mean constraints

$$\mu_{(k)}(\boldsymbol{X}_{(k)j}^T \boldsymbol{\beta}_{(k)}) = \sum_{i=1}^{n} p_i Y_{(k)i} \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\} \quad j = 1, \dots, n, k = 1, \dots, K$$

Table of Contents

Code

Currently the model is fit computationally in MATLAB using non-linear constrained optimization to maximise the empirical log-likelihood. [2]

[2] The code is available at github.

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i
- nK tilt parameters $\boldsymbol{\theta}_i$

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i
- nK tilt parameters $\boldsymbol{\theta}_i$

which are subject to $n(K + 1)$ mean and normalising constraints.

Table of Contents

Motivating Example (Butterfly Dataset)

This dataset contains 66 counts of 14 species of butterflies in Boulder Colorado, USA.

Table 2: Snippet of Butterfly Counts for the 3 most common species (Hui et al. 2013; Oliver, Prudic and Collinge 2006)

site	Pieris.rapae	Colias.philodice	Colias.eurytheme	building	vegetation	habitat
1	0	0	0	2.1	10.8	mixed
2	1	1	2	2.1	10.8	mixed
3	1	1	2	19.8	1.7	mixed
4	0	2	1	5.3	0	mixed
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Butterfly Dataset

We can fit this model directly to the counts of the each butterfly species using separate mean models, without the need to specify any association between species.

Butterfly Dataset

We can fit this model directly to the counts of the each butterfly species using separate mean models, without the need to specify any association between species.

$$\mu_{(k)i} = \mathbb{E}[Y_{(k)i}|\boldsymbol{X}_i] = \exp\{\boldsymbol{X}_i^T\boldsymbol{\beta}_{(k)}\},$$
$$i = 1, 2, \ldots, 66, \quad k = 1, 2 \ldots, 14$$

Butterfly Dataset

```
butterfly = fit_vspglm(...
         ["colias.philodice ~ (building,urban,habitat)",...
         "pieris.rapae ~ (building, urban, habitat)", ...
         "colias.eurytheme ~ (building, urban, habitat)"], ...
         tbl, {'log', 'log', 'log'})
```

Figure 1: Formula Syntax to fit 3 separate unconstrained models using the same covariates and log link functions

Butterfly Dataset

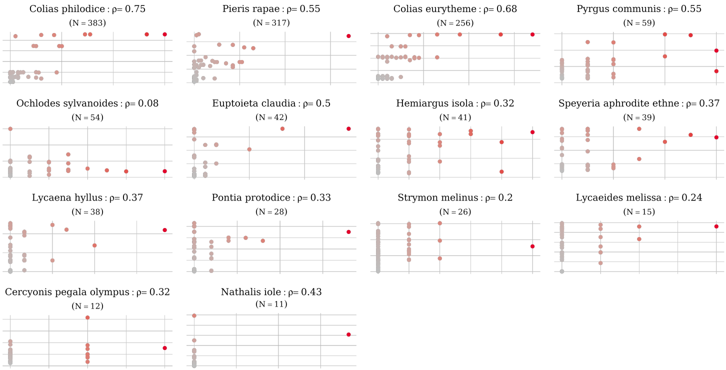

Figure 2: Predictions (y-axis) vs Observed counts (x-axis)

Butterfly Dataset

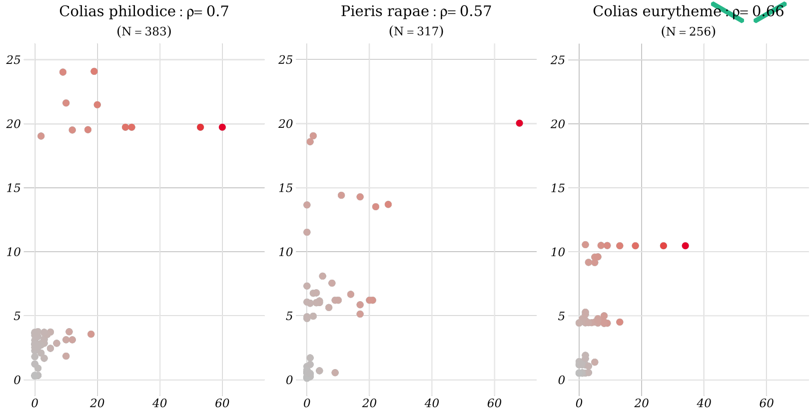

Figure 3: Predictions (y-axis) vs Observed counts (x-axis) for the 3 butterfly species with the greatest number of observed counts

Sorbinil Dataset

Contains $n = 41$ observations of itching scores (Y_L, Y_R) for the left and right eye after the application of sorbinil or a placebo, measured on a 9 point Likert scale between 0 and 4 with increments of 0.5.

Table 3: Snippet of Sorbinil dataset (Rosner, Glynn and Lee 2006)

$n = 6$		$n = 14$		$n = 14$		$n = 7$	
Sorbinil Left	**Sorbinil Right**	**Sorbinil Left**	**Placebo Right**	**Placebo Left**	**Sorbinil Right**	**Placebo Left**	**Placebo Right**
2	2	1	1.5	2.5	2	3	3
1	1	2	2.5	2.5	2.5	2	3
0.5	2	3	1	3	3	2.5	2.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0,1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

This model finds a significant effect that sorbinil reduces itching scores by 0.43.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

This model finds a significant effect that sorbinil reduces itching scores by 0.43.

The code to fit this symmetric model is

```
sorbinil = fit_vspglm(["(Y_L, Y_R) ~ (I_L & I_R)"], tbl, {'id'})
```

Sorbinil Dataset

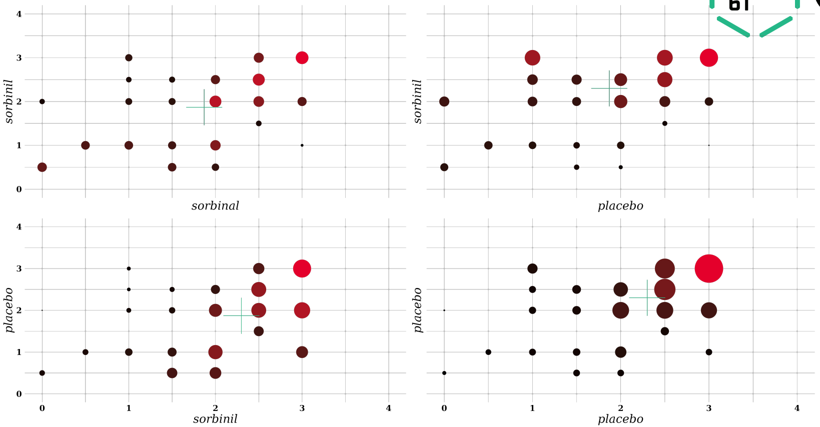

Figure 4: Reweighted Sorbinal pmf at group medians (left eye = y axis, right eye = x axis). Medians are indicated by the green cross. Areas with greater probability mass colored in vermillion.

Bibliography I

Huang, Alan (2014). "Joint estimation of the mean and error distribution in generalized linear models". In: *Journal of the American Statistical Association* 109.505, pp. 186–196.

Huang, Alan and Paul Rathouz (2012). "Proportional likelihood ratio models for mean regression". In: *Biometrika* 99.1, pp. 223–229.

Hui, Francis K. et al. (2013). "To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models". In: *Ecology* 94.9, pp. 1913–1919.

Oliver, J. C., K. L. Prudic and S. K. Collinge (2006). "Boulder County Open Space Butterfly Diversity and Abundance: "Ecological Archives" E087-061". In: *Ecology (Durham)* 87.4, p. 1066.

Rathouz, Paul J. and Liping Gao (2009). "Generalized linear models with unspecified reference distribution". In: *Biostatistics* 10.2, pp. 205–218.

Rosner, Bernard, Robert J Glynn and Mei-Ling T Lee (2006). "Extension of the rank sum test for clustered data: Two-group comparisons with group membership defined at the subunit level". In: *Biometrics* 62.4, pp. 1251–1259.

Trivariate Poisson Model

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, \sigma_0^2), \quad \sigma \in \mathbb{R}_+ \\
\boldsymbol{X} &\sim \ \mathrm{U}[-1, 1]^3 \\
Y_1 | \alpha, \boldsymbol{X}_{(1)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(1)}^T \boldsymbol{\beta}_{(1)} + \alpha)) \\
Y_2 | \alpha, \boldsymbol{X}_{(2)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(2)}^T \boldsymbol{\beta}_{(2)} + 0.5\alpha)) \\
Y_3 | \alpha, \boldsymbol{X}_{(3)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(3)}^T \boldsymbol{\beta}_{(3)} - 0.3\alpha))
\end{aligned}
$$

Trivariate Poisson Simulation Results

Table 4: Simulation results for trivariate Poisson model using a sample size of $n = 200$ and $N = 1000$ simulations

β	$\hat{\beta}$	$\lvert\beta - \hat{\beta}\rvert$	Errors		$p \leqslant 0.05$	CI		
			$\hat{\sigma}$	$\bar{se}(\hat{\beta})$		90%	95%	99%
0.4	0.40	0.004	0.13	0.12	0.88	0.89	0.94	0.99
-0.8	-0.79	0.002	0.13	0.14	0.99	0.93	0.96	0.99
0	0.001	0.001	0.13	0.12	0.05	0.89	0.95	0.99

Trivariate Mixed Effects Simulation

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0,\ \sigma_0^2) \\
\boldsymbol{X} &\sim \mathrm{U}[-1, 1]^3 \\
\boldsymbol{Y}_1 | \boldsymbol{X}_{(1)}, \alpha &\sim \mathcal{N}(\boldsymbol{X}_{(1)}^T \boldsymbol{\beta}_{(1)} + \alpha,\ \sigma_1^2) \\
\boldsymbol{Y}_2 | \boldsymbol{X}_{(2)}, \alpha &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(2)}^T \boldsymbol{\beta}_{(2)} + 0.5\alpha)) \\
\boldsymbol{Y}_3 | \boldsymbol{X}_{(3)}, \alpha &\sim \mathrm{Gamma}(\lambda,\ \exp(\boldsymbol{X}_{(3)}^T \boldsymbol{\beta}_{(3)} - 0.3\alpha))
\end{aligned}
$$

Trivariate Mixed Effects Simulation Results

Table 5: Simulation results for trivariate mixed effects model using a sample size of $n = 200$ and $N = 1000$ simulations

Margin	β	$\hat{\beta}$	$\lvert\beta - \hat{\beta}\rvert$	Errors		$p \leqslant 0.05$	CI		
				$\hat{\sigma}$	$\bar{\text{se}}(\hat{\beta})$		90%	95%	99%
Normal	1	1.005	0.0058	0.17	0.18	1	0.90	0.95	0.99
Poisson	-0.5	-0.49	0.0005	0.13	0.13	0.96	0.83	0.88	0.93
Gamma	0.4	0.39	0.004	0.12	0.13	0.85	0.89	0.93	0.97

Multivariate Normal Simulation Results

Table 6: Simulation results for bivariate normal model using sample size of $n = 200$ and $N = 1000$ simulations

β	$\hat{\beta}$	$\lvert\beta - \hat{\beta}\rvert$	Errors		$p \leqslant 0.05$	CI		
			$\hat{\sigma}$	$\bar{\text{se}}(\hat{\beta})$		90%	95%	99%
-1	-0.99	0.0004	0.11	0.11	1	0.91	0.95	0.99
0	-0.0004	0.0004	0.11	0.11	0.05	0.90	0.95	0.99
0.5	0.49	0.009	0.14	0.13	0.95	0.88	0.94	0.98
2.2	2.19	0.003	0.14	0.14	1	0.90	0.94	0.98