A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures

Martial Luyts, Geert Molenberghs, Geert Verbeke, Koen Matthijs, Eduardo E. Ribeiro Jr, Clarice G. B. Demétrio, and John Hinde

Supplementary Materials

A General overview of over- and/or underdispersed count models

Due to the restricted mean-variance relationship of the Poisson log-linear GLM models (i.e., $\text{E}(Y_i) = \text{Var}(Y_i) \equiv \lambda_i$, where $Y_i \in \mathbb{N}$, $i = 1, \ldots, n$, is Poisson distributed with parameter $\lambda_i \in \mathbb{R}_{>0}$, the set of positive real numbers, and with $\ln(\lambda_i) = \mathbf{x}_i' \cdot \boldsymbol{\beta}$, where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ denote a p-dimensional vector of covariates for ith observation and the associated parameter vector, respectively), various extensions have been proposed in the literature (Breslow, 1984; Lawless, 1987; Hinde and Demétrio, 1998) - and in what follows, some popular models will be examined in more detail.

1 Quasi-Poisson model

A straightforward modification, in the context of the exponential family, is to allow the dispersion (scale) parameter, denoted by δ, to not be restricted to 1. This leads us to $\mathrm{Var}(Y_i) = \delta \cdot \mathrm{E}(Y_i)$, where $\delta > 1$ and $\delta < 1$ indicates over- and underdispersion, respectively. This results in the so-called quasi-Poisson model (Wedderburn, 1974), where the point estimates of $\boldsymbol{\beta}$ are identical to those of the standard Poisson model, but standard errors are scaled by $\sqrt{\delta}$ resulting in possible differences in inferences on covariates compared to the standard Poisson model.

2 Negative binomial model

Another elegant way to provide flexibility is through a two-stage model. A popular approach in this context is to assume that $Y_i \mid \lambda_i \sim \mathrm{Poi}(\lambda_i)$ and that the parameter λ_i is itself a random variable with mean μ_i and variance σ_i^2. By using standard results on iterated expectations we have:

$$\mathrm{E}(Y_i) = \mu_i, \qquad \mathrm{Var}(Y_i) = \mu_i + \sigma_i^2.$$

A popular specific distributional choice is $\lambda_i \sim \mathrm{Gamma}(\alpha, \alpha^{-1})$, for reasons of identiability (Duchateau and Janssen, 2007), leading to the negative binomial (NB) model. Choosing the Gamma distribution has the advantage of (1) satisfying the mean's scale for count outcomes and (2) obtaining closed forms for the marginal mean and variance, and even for the entire marginal distribution (Molenberghs et al., 2007). The corresponding (marginal) probability mass function, mean and variance of the model are equal to

$$P(Y_i = y_i \mid \mathbf{x}_i) = \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\cdot\Gamma(\alpha^{-1})} \cdot \left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda_i}\right)^{\alpha^{-1}} \cdot \left(\frac{\lambda_i}{\alpha^{-1}+\lambda_i}\right)^{y}, \text{ with } \lambda_i = \mathrm{e}^{\mathbf{x}_i'\cdot\beta},$$
$$\mathrm{E}(Y_i) = \lambda_i, \qquad \mathrm{Var}(Y_i) = \lambda_i + \alpha \cdot \lambda_i^2,$$

respectively (Lawless, 1987; Cameron and Trivedi, 1986). We should note that, from a hierarchical/conditional viewpoint, only overdispersion can be examined (since for a valid Gamma distribution $\alpha > 0$).

3 Conway-Maxwell-Poisson model

The Conway-Maxwell-Poisson (COM) model, first introduced by Conway and Maxwell (1962), is suitable for analysing count data that exhibit either over- or underdispersion. Even though its existence has been known for several decades, most research on this model was done during the last decade. Shmueli et al. (2005), for example, investigated the statistical properties of the COM distribution. While in a Bayesian context, Kadane et al. (2006) developed the conjugate distributions for the parameters of the COM distribution. The probability mass function of the model can be expressed as

$$P(Y_i = y_i \mid \mathbf{x}_i) = \frac{1}{Z(\lambda_i,\tau)} \cdot \frac{\lambda_i^{y_i}}{(y_i!)^\tau},$$
$$\text{with } \lambda_i = \mathrm{e}^{\mathbf{x}_i'\cdot\boldsymbol{\beta}}, \qquad Z(\lambda_i,\tau) = \sum_{n=0}^{+\infty} \frac{\lambda_i^n}{(n!)^\tau}.$$

The domain of admissible parameters for which the probability mass function above defines a probability distribution is $(\lambda_i, \tau) > 0$, and $0 < \lambda_i < 1, \tau = 0$. Some well-known discrete data models result from this. When τ equals 1, it reduces to the standard Poisson model. When $\tau \to +\infty$, the COM model approaches a Bernoulli model with success parameter $\pi_i = \frac{\lambda_i}{1+\lambda_i}$. While if $\tau = 0$ and $\lambda_i < 1$, the geometric model with success probability $1-\lambda_i$ is obtained. In terms of dispersion, specific focus is put on the mean and variance functions and the nature of the different dispersion regions can be found in Appendix E2.

The mean and variance can be approximated by

$$\mathrm{E}(Y_i) = \lambda_i \cdot \frac{\partial \mathrm{log} Z(\lambda_i,\tau)}{\partial\lambda_i} \approx \lambda_i^{1/\tau} - \frac{\tau-1}{2\cdot\tau}, \qquad \mathrm{Var}(Y_i) = \frac{\partial \mathrm{E}(Y_i)}{\partial \mathrm{log}\lambda_i} \approx \frac{1}{\tau}\cdot\lambda_i^{1/\tau}.$$

4 Double Poisson model

The double Poisson (DP) model, based on the double-exponential family of Efron (1986), has hardly been investigated and applied since its first introduction three decades ago. Winkelmann (2008) and Hilbe (2011) indicated that the normalizing constant is the bottleneck in applying the DP by showing that fitted models with its normalizing constant approximated by Efron's original method are not exact. For these and other reasons, different approximations have been proposed in the literature. A full discussion can be found in Zou et al. (2013).

The probability mass function of the DP model can be written as

$$P(Y_i = y_i \mid \mathbf{x}_i) = K(\lambda_i,\phi)\cdot\phi^{1/2}\cdot \mathrm{e}^{-\phi\cdot\lambda_i}\cdot\frac{\mathrm{e}^{-y_i}\cdot y_i^{y_i}}{y_i!}\cdot\left(\frac{\mathrm{e}\cdot\lambda_i}{y_i}\right)^{\phi\cdot y_i},$$
$$\lambda_i = \mathrm{e}^{\mathbf{x}_i'\cdot\boldsymbol{\beta}}, \qquad \frac{1}{K(\lambda_i,\phi)} \approx 1 + \frac{1-\phi}{12\cdot\phi\cdot\lambda_i}\cdot\left(1+\frac{1}{\phi\cdot\lambda_i}\right),$$

where $K(\lambda_i,\phi)$ is the normalizing constant that is often close to 1. The corresponding mean and variance can be approximated by

$$\mathrm{E}(Y_i) \approx \lambda_i, \qquad \mathrm{Var}(Y_i) \approx \frac{\lambda_i}{\phi}.$$

Thus, the DP model allows for both overdispersion ($\phi < 1$) and underdispersion ($\phi > 1$). While for $\phi = 1$, the standard Poisson model results.

B Proof of dispersion for DE case

Theorem 1. *Let Y_i, $i = 1, \ldots, n$, be (type 1) DE distributed, with $\lambda > 0$. Then, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$).*

Proof. To prove it, three situations are examined:

1. $\underline{\mathrm{E}(Y_i) > \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})} > \frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} < 0 \qquad \Rightarrow \perp$$

2. $\underline{\mathrm{E}(Y_i) = \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})} = \frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} = 0 \qquad \Rightarrow \lambda \to +\infty$$

3. $\underline{\mathrm{E}(Y_i) < \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})} < \frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} > 0 \qquad \Rightarrow \surd$$

   Thus, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$)!

□

C Proof of mean and variance convergence for DW case

Lemma 2 (d'Alembert's ratio test). *Let $\sum_{n=0}^{+\infty} a_n$ be an infinite serie, and consider*

$$L = \lim_{n \to +\infty} \mid \frac{a_{n+1}}{a_n} \mid .$$

1. *If $L < 1$, then the series converges absolutely;*
2. *If $L > 1$, then the series diverges;*
3. *If $L = 1$ or the limit fails to exist, then the test is inconclusive.*

Lemma 3 (Raabe–Duhamel's test). *Let $a_n > 0$ $(\forall n)$. Define*

$$b_n = n \cdot \left(\frac{a_n}{a_{n+1}} - 1 \right).$$

If $L = \lim_{n \to +\infty} b_n$ exists, there are three possibilities:

1. *If $L > 1$, then the series $\sum_{n=0}^{+\infty} a_n$ converges;*
2. *If $L < 1$, then the series $\sum_{n=0}^{+\infty} a_n$ diverges;*
3. *If $L = 1$, then the test is inconclusive.*

Theorem 4. *Let Y_i, $i = 1, \ldots, n$, be (type 1) DW distributed. Then, it can be shown that*

$$\begin{aligned} \mathrm{E}(Y_i) &\left(= \mu = \sum_{n=1}^{+\infty} q^{n^\rho} \right) < +\infty, \\ \mathrm{Var}(Y_i) &\left(= 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^{n^\rho} - \mu - \mu^2 \right) < +\infty. \end{aligned}$$

Proof. A trivial proof can be conducted for $\rho \geq 1$, since $\sum_{n=0}^{+\infty} q^{n^\rho} \leq \sum_{n=0}^{+\infty} q^n = (1-q)^{-1}$, $\sum_{n=0}^{+\infty} n \cdot q^{n^\rho} \leq \sum_{n=0}^{+\infty} n \cdot q^n$, and the series $\sum_{n=0}^{+\infty} n \cdot q^n$ converges. Indeed, using Lemma 2 with $a_n = n \cdot q^n$, it can easily be shown that $L < 1$. For $\rho < 1$, Lemma 3 can be used, where $a_n = n \cdot [q^{n^\rho} - q^{(n+1)^\rho}]$ and $a_n = n^2 \cdot [q^{n^\rho} - q^{(n+1)^\rho}]$ are proper choices for $\mathrm{E}(Y_i)$ and $\mathrm{Var}(Y_i)$, respectively. $\square$

Additionally, based on the integral test (Knopp, 1951) and assuming $q = e^{-\lambda}$, the following lower and upper boundaries can be obtained for the mean and variance expression $(\forall t \in 1, 2, \ldots)$:

$$\begin{aligned}\mathrm{E}(Y_i) &\in \left[\sum_{n=1}^{t} q^{n^\rho} + \int_{t+1}^{+\infty} q^{n^\rho} \cdot dn; \sum_{n=1}^{t} q^{n^\rho} + \int_{t}^{+\infty} q^{n^\rho} \cdot dn\right] \\ &\in \left[\sum_{n=1}^{t} e^{-\lambda \cdot n^\rho} + \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t+1)^\rho]; \sum_{n=1}^{t} e^{-\lambda \cdot n^\rho} + \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho)\right],\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Y_i) \in \Bigg[& 2 \cdot \sum_{n=1}^{t} n \cdot q^{n^\rho} - \sum_{n=1}^{t} q^{n^\rho} - \left(\sum_{n=1}^{t} q^{n^\rho}\right)^2 - \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho) \cdot \sum_{n=1}^{t} q^{n^\rho} - \frac{1}{\rho^2 \cdot \lambda^{2/\rho}} \\ & \cdot [\Gamma(1/\rho; \lambda \cdot t^\rho)]^2 - \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho) + \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[2/\rho; \lambda \cdot (t+1)^\rho]; \\ & 2 \cdot \sum_{n=1}^{t} n \cdot q^{n^\rho} - \sum_{n=1}^{t} q^{n^\rho} - \left(\sum_{n=1}^{t} q^{n^\rho}\right)^2 - \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t+1)^\rho] \cdot \sum_{n=1}^{t} q^{n^\rho} \\ & - \frac{1}{\rho^2 \cdot \lambda^{2/\rho}} \cdot \{\Gamma[1/\rho; \lambda \cdot (t+1)^\rho]\}^2 - \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t+1)^\rho] + \frac{2}{\rho \cdot \lambda^{1/\rho}} \\ & \cdot \Gamma(2/\rho; \lambda \cdot t^\rho) \Bigg].\end{aligned}$$

D SAS code for the hierarchical DE and DW model

```
/***************************
SOFTWARE: SAS 9.4.
OBJECTIVE: Analyzing Moerzeke data with the DE and DW approach;
DATASET: Moerzeke data, containing information about 457 families;
VARIABLE DESCRIPTION:
- ID: Family ID;
- FamilyMember: Family member indicator, i.e., F = father,
        M = mother, C = first born child;
- Sexnum: Binary indicator of the gender of first born child, i.e.,
        1 = boy, 0 = girl;
- y: Discretised life expectancy of a household member;
AUTHOR: M. Luyts (L-Biostat);
****************************/

libname m 'C:\Users\u0106491\Desktop\Moerzeke data';
```

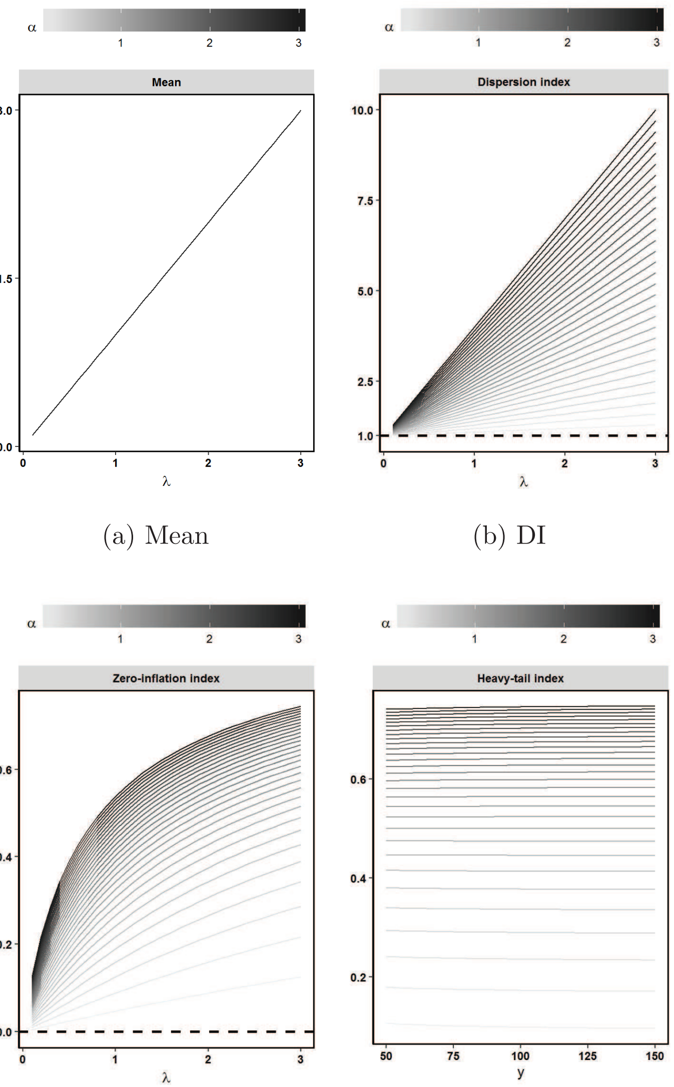
```
/* Model from exponential case, via NLMIXED */
proc nlmixed data = m.MoerzekeFinal;
  parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
          beta1X=-0.018 beta1XX=-0.018 sigma=1;
  if FamilyMember='F' then
     eta=beta0XX + beta1XX*SexNum + u;
  else if FamilyMember='M' then
     eta=beta0X + beta1X*SexNum + u;
  else eta=beta0 + beta1*SexNum + u;
  expeta = exp(eta);
  ll = eta*y - log(expeta + 1)*y + log(1 - (expeta/(expeta+1)));
  model y ~ general(ll);
  random u ~ normal(0, exp(sigma)**2) subject=id;
  estimate 'random intercept' exp(sigma);
run;

/* Model from Weibull case, via NLMIXED */
proc nlmixed data = m.MoerzekeFinal;
  parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
          beta1X=-0.018 beta1XX=-0.018 sigma=1 rho=1;
  if FamilyMember='F' then
     eta=beta0XX + beta1XX*SexNum + u;
  else if FamilyMember='M' then
     eta=beta0X + beta1X*SexNum + u;
  else eta=beta0 + beta1*SexNum + u;
  lambda = log(exp(eta)+1) - log(exp(eta));
  if y=0 then prob = 1 - exp(-1*lambda);
  else prob=exp(-1*lambda*(y**rho))-exp(-1*lambda*((y+1)**rho));
  ll = log(prob);
  model y ~ general(ll);
  random u ~ normal(0, exp(sigma)**2) subject=id;
  estimate 'random intercept' exp(sigma);
run;
```

E Characteristic indices for the COM, DP and NB

In spirit of Section 3, similar characteristic indices w.r.t. the Poisson distribution can be obtained for the COM, DP and NB distributions.

1 Negative binomial distribution

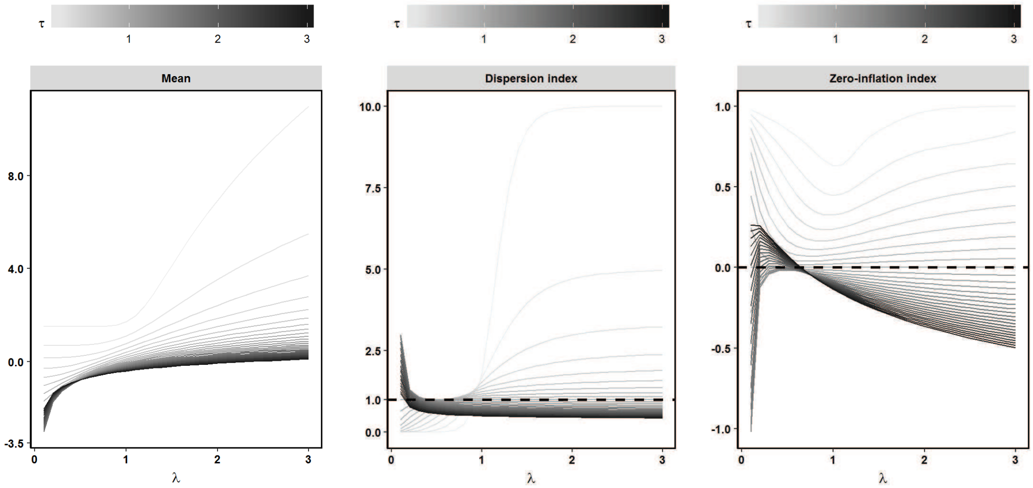


(a) Mean (b) DI

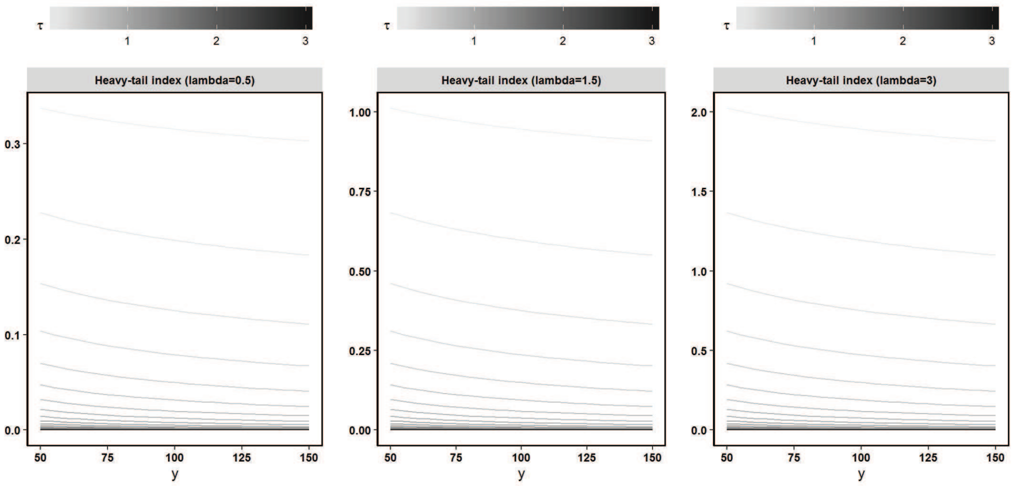
(c) ZI (d) HT

Figure 7: *Characteristic indices of NB related to the Poisson distribution.*

2 Conway-Maxwell-Poisson distribution



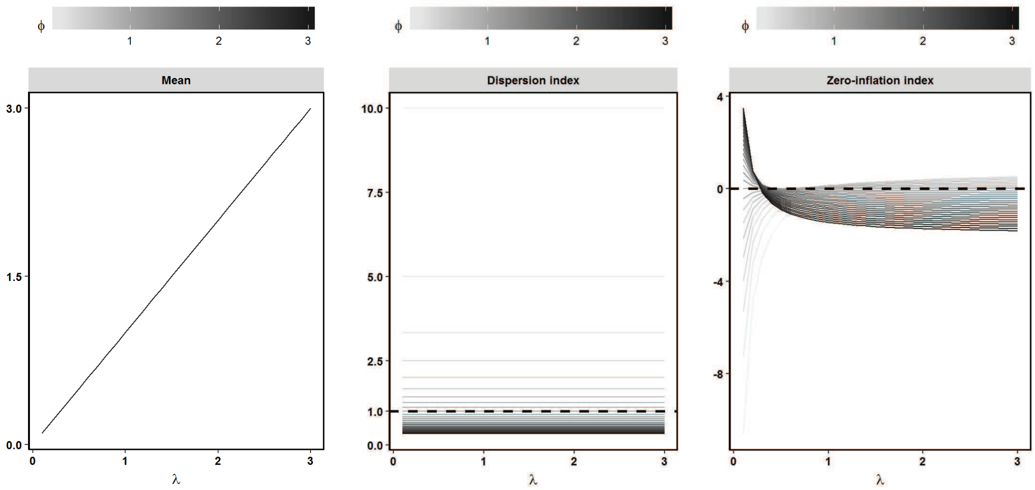
(a) Mean (log-scale) (b) DI (c) ZI



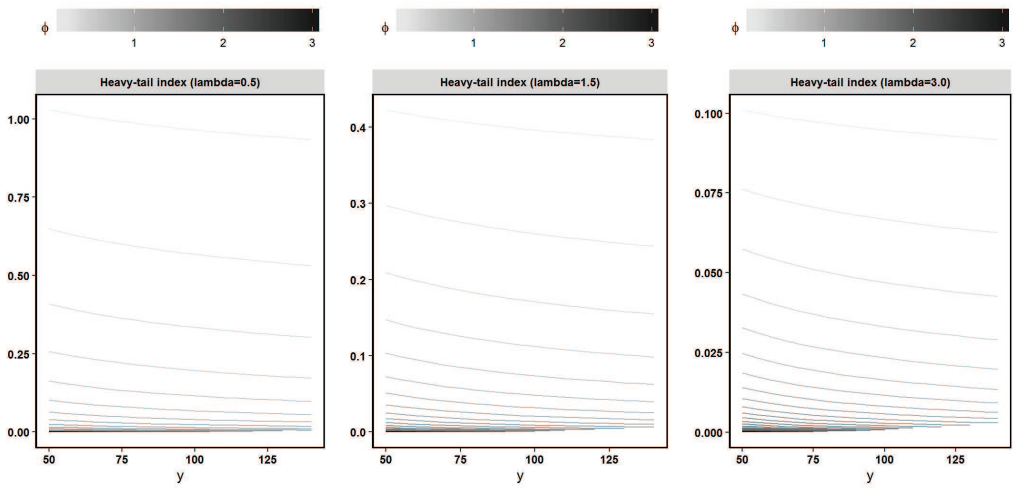
(d) HT for varying λ

Figure 8: *Characteristic indices of COM related to the Poisson distribution.*

3 Double Poisson distribution



(a) Mean (b) DI (c) ZI



(d) HT for varying λ

Figure 9: *Characteristic indices of DP related to the Poisson distribution.*