**Ricardo R. Petterle[1,2] / Wagner H. Bonat[3] / Célestin C. Kokonendji[4] / Juliane C. Seganfredo[5] / Atamai Moraes[5] / Monica G. da Silva[5]**

# Double Poisson-Tweedie Regression Models

## Omitted

[1] Sector of Health Sciences, Medical School, Paraná Federal University, Curitiba, Brazil
[2] Sector of Health Sciences, Medical School, Universidade Federal do Parana, Curitiba, Brazil
[3] Department of Statistics, Paraná Federal University, Curitiba, Brazil, E-mail: wbonat@ufpr.br. https://orcid.org/0000-0002-0349-7054.
[4] Laboratoire de Mathématiques de Besançon, Bourgogne Franche-Comté University, Besançon, France
[5] Departamento de Saúde Sanitária, Paraná Federal University, Curitiba, Brazil

**Abstract:**
In this paper, we further extend the recently proposed Poisson-Tweedie regression models to include a linear predictor for the dispersion as well as for the expectation of the count response variable. The family of the considered models is specified using only second-moments assumptions, where the variance of the count response has the form $\mu + \phi\mu^p$, where $\mu$ is the expectation, $\varphi$ and $p$ are the dispersion and power parameters, respectively. Parameter estimations are carried out using an estimating function approach obtained by combining the quasi-score and Pearson estimating functions. The performance of the fitting algorithm is investigated through simulation studies. The results showed that our estimating function approach provides consistent estimators for both mean and dispersion parameters. The class of models is motivated by a data set concerning CD4 counting in HIV-positive pregnant women assisted in a public hospital in Curitiba, Paraná, Brazil. Specifically, we investigate the effects of a set of covariates in both expectation and dispersion structures. Our results showed that women living out of the capital Curitiba, with viral load equal or larger than 1000 copies and with previous diagnostic of HIV infection, present lower levels of CD4 cell count. Furthermore, we detected that the time to initiate the antiretroviral therapy decreases the data dispersion. The data set and R code are available as supplementary materials.

**Keywords:** Poisson-Tweedie distribution, double generalized linear models, estimating functions, CD4 count, human immunodeficiency virus (HIV), overdispersion

**DOI:** 10.1515/ijb-2018-0119

**Received:** November 14, 2018; **Revised:** March 18, 2019; **Accepted:** April 2, 2019

## 1 Introduction

Acquired immunodeficiency syndrome (AIDS) is one of the main public health problems at worldwide levels [1]. Nowadays, around 36.7 million of people are (human immunodeficiency virus) HIV-positive around the world and among them 17.8 million are women [2]. In spite of the recent developments in the treatment of HIV-positive patients, keeping HIV controlled is a challenging task. This could be even more challenging for pregnant women, since to avoid mother-to-child transmission it is necessary to achieve undetectable levels of viral load [3]. Thus, early access to prenatal care and early initiation of antiretroviral therapy (ART) is required. Keeping the pregnant in care since early prenatal favors the ART adaptation, ensuring tolerance, the success of treatment and thus preventing the vertical transmission [4]. The CD4 cell count is an important marker to evaluate the immunologic system and the progression of human immunodeficiency infections of HIV-positive patients. Laboratorial AIDS diagnosis is confirmed when CD4 cell count is under 200 cell/mm$^3$ [5]. Consequently, these patients have an increased risk of death due to opportunistic infections [6].

Recently, some authors have turned their attention to study factors associated with CD4 cell count in HIV-positive patients. Grover et al. [5] compared the fit of Poisson, negative binomial and generalized Poisson regression models to investigate the effect of various socio-demographic covariates such as age, gender, geographical location and drug usage in CD4 count of AIDS patients in India. Similarly, Seyoum and Zewotir [7] compared the fit of the quasi-Poisson and negative binomial regression models to identify factors associated with CD4 count in adults at the beginning of antiretroviral treatment in North-West Ethiopia. Helleberg et al. [8] analyzed

**Wagner H. Bonat** is the corresponding author.

data from the Danish nationwide, population-based cohort study in the period 1995–2000 with quarterly CD4 measurements and detected that the risk of cardiovascular disease, cancer and death increased markedly after CD4 count declines. In this article, we aim to investigate factors associated with the improvement in CD4 cell count of HIV-positive pregnant women assisted by a public hospital in Curitiba, Paraná, Brazil. Thus, suitable regression models for count data are demanded for such an investigation.

In the context of count regression models, the Poisson model is the most popular one. However, it is well-known that such a regression model is limited, since it assumes equidispersion, i.e. mean equals the variance [9]. In practical data analysis, we can have both under/overdispersed counts being the former more often found. One possible cause of under/overdispersion is departure from the Poisson process. The Poisson counts can be interpreted as the number of events in a given time interval where the arrivals times are exponential distributed. In the cases where this assumption is violated the resulting counts can be under or overdispersed [10]. Another possibility and probably more frequent cause of overdispersion is unobserved heterogeneity of experimental units. It can be due, for example, to correlation between individual responses, cluster sampling, omitted covariates and others. These departures from the Poisson distribution are manifested in the raw data as a zero-inflated or heavy-tailed count distribution. The consequences of failing to take into account the under or overdispersion in the analysis of count data are distinct. In the case of overdispersion the standard errors associated with the regression coefficients computed under the equidispersion assumption are too optimistic and associated hypothesis tests will tend to give false positive results by incorrectly rejecting null hypotheses. The opposite situation will happen in the case of underdispersion. In both cases, the Poisson model provides unreliable standard errors for the regression coefficients and hence potentially misleading inferences.

The statistical literature for dealing with overdispersed count data has grown quickly in the last years, see for example [11–17] and references therein. Over all these approaches, probably the class of extended Poisson-Tweedie regression models is the most flexible, since it has as special cases the Hermite, Neyman Type A, Pólya Aeppli, negative binomial and Poisson inverse Gaussian regression models. Furthermore, the extended Poisson-Tweedie model can mimic the behavior of other count regression models as the Gamma-Count [10] and COM-Poisson [18]. In spite of their flexibility the Poisson-Tweedie regression models allow modeling only the expectation of the count response variable as a function of the covariates. Smyth [19] argued that it is quite common for data sets to show evidence of systematic variation in the dispersion structure. Furthermore, the author showed that the correct modeling of the dispersion heterogeneity always reduces the standard error of the mean regression coefficients, which in turn provides a genuine increase in precision. Thus, we claim that it is very useful and attractive to have the possibility to detect and model dispersion heterogeneity in the context of Poisson-Tweedie regression models.

The main goal of this article is to further extend the Poisson-Tweedie regression models recently proposed by [11]. We use the principles of the double generalized linear models [19, 20], where we model the mean and dispersion structures by means of a link function and a linear predictor. The corresponding class of double Poisson-Tweedie regression models is specified using only second-moments assumptions and can easily be fitted using the estimating function approach proposed in [21]. It is interesting to note that the approach used for estimation and inference resembles Wedderburn's quasi-likelihood [22] and consequently one could name our model double quasi Poisson-Tweedie regression models. However, since the term quasi would here only refer to an aspect of the estimation routine rather than to a model property, we opted to omit it from the title. The proposed model is exemplified by a data set concerning CD4 counting in HIV-positive pregnant women assisted in a public hospital in Curitiba, Paraná, Brazil.

Section 2 describes the data set. In the Section 3 we motivate and present the double Poisson-Tweedie regression models. Estimation and inference for the proposed models are presented in Section 4. The main results from two comprehensive simulation studies are described in the Section 5. In the Section 6 we apply the double Poisson-Tweedie regression models to investigate factors associated with CD4 count in HIV-positive pregnant women. The results are discussed in Section 7, including some directions for future investigations. Finally, the `R` code, data set and some additional Figures are presented in the supplementary material web page http://www.leg.ufpr.br/doku.php/publications:papercompanions:dptw.

## 2 Data set

The data set used here was obtained from the Hospital de Clínicas de Curitiba of the University Federal of Paraná (HC-UFPR). The data were collected from hospital's database and comprise of a cohort of 379 HIV-positive pregnant women who had at least one prenatal appointment which gave birth at the hospital from February 2011 to December 2015. Demographic, clinical and laboratory data were obtained from review of medical files.

The main goal of this study is to investigate the effect of a set of covariates in the CD4 counts. The covariates are: `GA` - gestational age in weeks, factor having two levels (0: $\leq$ 27; 1: $\geq$ 28); `HIVD` - HIV diagnostic during pregnancy, factor having two levels (0: No; 1: Yes); `CR` - city of residence, factor having two levels (0: Curitiba; 1: Other city); `VL` - viral load, factor having three levels (0: Undetectable; 1: $\leq$ 999; 2: $\geq$ 1000); `CO` - co-infection, factor having two levels (0: No; 1: Yes); `NBC` - newborn condition, factor having three levels (0: HIV-negative; 1: HIV-positive; 2: Unknown); `HIVRF` - known HIV risk factors, factor having three levels (0: Unknown; 1: Drugs; 2: Vertical); `DEL` - delivery, factor having two levels (0: Vaginal; 1: Caesarian); `URG` - use of raltegravir, factor having two levels (0: No; 1: Yes); `TART` - how long the patient has been on antiretroviral therapy (ART) in weeks; `GAART` - gestational age at which the patient started ART and `AGE` - patient age.

Figure 1 presents dispersion diagrams and boxplots to investigate the association of the CD4 count with the set of covariates. The plots suggest that the covariates HIV diagnostic during pregnancy (Figure 1E), viral load(Figure 1G), known HIV risk factors (Figure 1J) and delivery (Figure 1K) are associated with the CD4 counts.

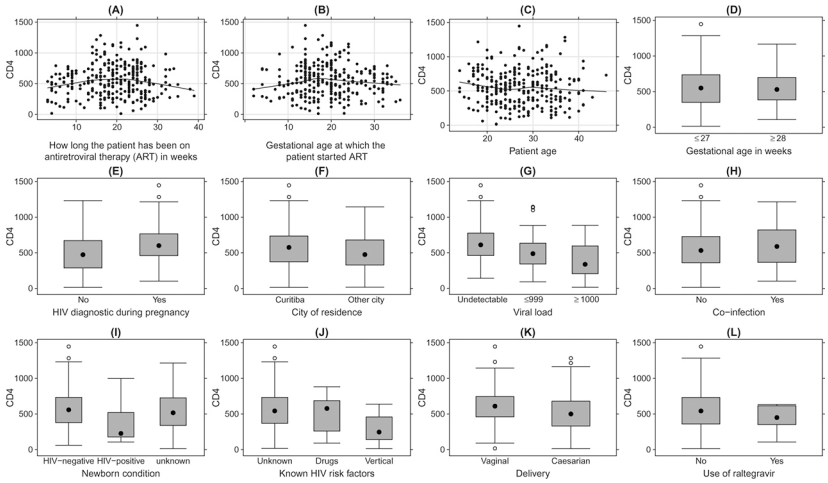


**Figure 1:** Dispersion diagrams (A to C) for the continuous covariates and boxplots (D to L) for the categorical covariates.

## 3 Double Poisson-Tweedie regression models

In this section, we motivate and propose the double Poisson-Tweedie regression models. The Poisson-Tweedie distributions for $p \geq 1$ are Poisson-Tweedie mixtures and consequently there is no closed-form expression for the probability mass function (pmf), for details see [12, 23, 24]. Bonat et al. [11] discussed the use of numerical methods to approximate the pmf of the Poisson-Tweedie distributions. The authors showed that the Monte Carlo method provides a reasonable approximation. Figure 2 presents the pmf for some Poisson-Tweedie distributions based on the Monte Carlo approximation. In all scenarios, the expectation $\mu$ was fixed at 10, however, we vary the values of the dispersion and power parameters to illustrate the flexibility of the distribution to deal with count data. The limiting case $p = 0$ corresponds to the Hermite distribution for which we have a closed-form expression available. Such a pmf is implemented in `R` through the `hermite` package [25].

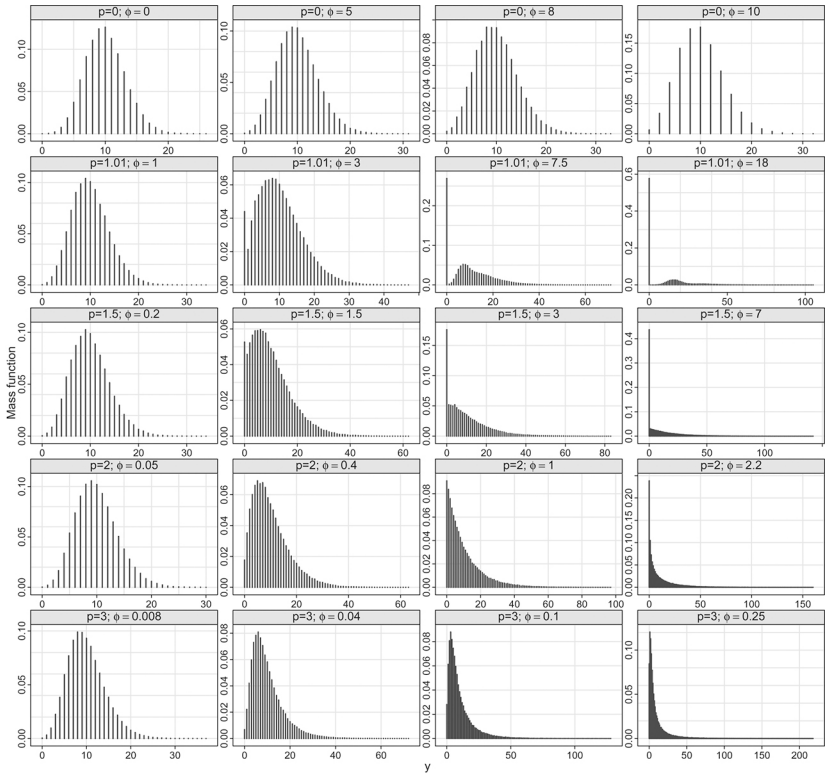


**Figure 2:** Probability mass function of Poisson-Tweedie distributions for different values of the dispersion ($\varphi$) and power ($p$) parameters.

Figure 2 shows that the shape of the Poisson-Tweedie distribution changes slightly for $p = 0$ when the dispersion parameter is increased. However, for larger values of the power parameter $p \geq 1$ the shape of the distribution is strongly determined by the dispersion parameter values. Thus, for small dispersion, we have shapes similar to the Poisson distribution (equidispersion). On the other hand, for large values of the dispersion parameter, we have zero-inflated and/or heavy-tailed count distributions. Thus, the pmf presented in Figure 2 highlight the importance to model the dispersion parameter, when dealing with overdispersed count data.

Despite of the pmf of the Poisson-Tweedie distribution is not available in closed-form, its first two moments (mean and variance) can easily be obtained. Jørgensen and Kokonendji [26] showed by using factorial cumulant generating functions that for $Y \sim PTw_p(\mu, \phi)$, $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu + \phi\mu^p$. This fact motivated [11] to specify a set of regression models based only on second-moments assumptions. In this article, we extend the approach proposed in [11] by regression of both the $\mu$ and $\varphi$ on the values of potential covariates. It is important to highlight that based only on second-moments assumptions the pmf of the Poisson-Tweedie distribution is not required for estimation and inference on double Poisson-Tweedie regression models. Furthermore, the restrictions of the power parameter space are no longer required.

Thus, consider a cross-section data set, $(y_i, \mathbf{x}_i)$, $i = 1,\ldots,n$, where $y_i$'s are independent and identically distributed (iid) realizations of $Y_i$ according to an unspecified count distribution, whose expectation and variance are given by

$$\begin{aligned} E(Y_i) &= \mu_i \\ Var(Y_i) &= \mu_i + \phi_i \mu_i^p, \end{aligned}$$

where $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ and $\phi_i = h^{-1}(\mathbf{z}_i^\top \boldsymbol{\gamma})$. In this notation, $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are $(S \times 1)$ vectors of known covariates and unknown regression coefficients associated to the expectation of the count response variable. Similarly, $\mathbf{z}_i$ and $\boldsymbol{\gamma}$ are $(Q \times 1)$ vectors of known covariates and unknown dispersion coefficients associated to the dispersion structure. Finally, both $g(\cdot)$ and $h(\cdot)$ are suitable link functions. In this article, we adopted the logarithm link function for both $g(\cdot)$ and $h(\cdot)$, since $\mu, \varphi > 0$.

We highlight that the power parameter brings more flexibility to the model being an index that distinguishes between some important distributions, such as the Hermite ($p = 0$), Neyman type A ($p = 1$), Pólya Aeppli ($p = 1.5$), negative binomial ($p = 2$) and Poisson inverse-Gaussian ($p = 3$). Furthermore, since our model is specified using only second-moments assumptions, the estimating function approach presented in Section 4 allows us to estimate the power parameter, which in turn works as an automatically model selector.

## 4 Estimation and inference

In this section, we shall present the estimating function approach adopted for parameter estimation and inference. Estimating functions could be explained as functions whose goal is to mimic the behavior and properties of the score function, i.e the first derivative of the log-likelihood function. In the context of maximum likelihood estimation the properties of the score function are key to define the asymptotic properties of the maximum likelihood estimators. Let $U(\theta|\boldsymbol{Y})$ denotes a score function, we can show that under regularity conditions [27] $E(U(\theta|\boldsymbol{Y})) = 0$ and $Var(U(\theta|\boldsymbol{Y})) = E(U^2(\theta|\boldsymbol{Y})) = E(-U'(\theta|\boldsymbol{Y}))$. These two results are combined using the delta method in order to obtain the asymptotic properties of the maximum likelihood estimators.

The score function is a special case of an estimating function. Note that the expectation of the score function is zero, it means that we have an unbiased estimating function. Furthermore, the computation of the variance and first derivative of the score function are key to apply the delta method. In the context of estimating functions the variance of an estimating function is called variability and the expectation of its first derivative is called sensitivity. Thus, it is clear that when using estimating functions their variability and sensitivity matrices are important to define the asymptotic properties of the estimating functions estimators.

It is important to highlight that in the context of maximum likelihood estimation the sensitivity and variability matrices coincide up to a sign resulting in the well-known Fisher information matrix. On the other hand, in the context of estimating functions the sensitivity and variability matrices do not always coincide, however, when they coincide we have an optimal estimating function. Thus, the choice of the estimating function and the computation of its sensitivity and variability matrices are the main interest when using an estimating function approach for parameter estimation and inference.

In order to present the estimating function approach adopted in this paper, we use the terminologies and results from [28], [21] and [11]. Our estimating function approach consists of combining the quasi-score and Pearson estimating functions for the estimation of the regression and dispersion parameters, respectively. The methodology is similar to the quasi-likelihood approach of [22].

The double Poisson-Tweedie regression models proposed in the Section 3 are described by two set of parameters, thus $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda} = (\boldsymbol{\gamma}, p)^\top)^\top$. Note that, $\boldsymbol{\lambda}$ is the $Q+1$ vector containing all parameters of the covariance structure.

For the estimation of the regression coefficients we adopted the quasi-score function given by

$$\psi_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\lambda}) = \left( \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_1} \sigma_i^{-1}(y_i - \mu_i), \dots, \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_S} \sigma_i^{-1}(y_i - \mu_i) \right)^\top, \tag{1}$$

where $\sigma_i = \mu_i + \phi_i \mu_i^p$ and $\partial \mu_i / \partial \beta_s = \mu_i x_{is}$ for $s = 1,\dots,S$.

The entry $(s, s')$ of the $S \times S$ sensitivity matrix of $\psi_{\boldsymbol{\beta}}$ is given by

$$\mathrm{S}_{\boldsymbol{\beta}_{ss'}} = \mathrm{E}\left( \frac{\partial}{\partial \beta_{s'}} \psi_{\beta_s}(\boldsymbol{\beta},\boldsymbol{\lambda}) \right) = -\sum_{i=1}^{n} \mu_i x_{is} \sigma_i^{-1} x_{is'} \mu_i. \tag{2}$$

Similarly, the entry $(s, s')$ of the $S \times S$ variability matrix of $\psi_{\boldsymbol{\beta}}$ has the form

$$\mathrm{V}_{\boldsymbol{\beta}_{ss'}} = \mathrm{Cov}(\psi_{\boldsymbol{\beta}_s}(\boldsymbol{\beta},\boldsymbol{\lambda}), \psi_{\boldsymbol{\beta}_{s'}}(\boldsymbol{\beta},\boldsymbol{\lambda})) = \sum_{i=1}^{n} \mu_i x_{is} \sigma_i^{-1} x_{is'} \mu_i. \tag{3}$$

We note in passing that the sensitivity and variability matrices in this case coincide up to a sign. Consequently, the quasi-score function is optimum for the estimation of the regression coefficients.

The dispersion parameters are estimated based on the following Pearson estimating function,

$$\psi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda},\boldsymbol{\beta}) = \left( \sum_{i=1}^{n} \omega_{i\lambda_1} \left[ (y_i - \mu_i)^2 - \sigma_i \right], \dots, \sum_{i=1}^{n} \omega_{i\lambda_{Q+1}} \left[ (y_i - \mu_i)^2 - \sigma_i \right] \right)^\top \tag{4}$$

where $\omega_{i\lambda_q} = -\partial \sigma_i^{-1} / \partial \lambda_q$ for $q = 1, \dots, Q+1$.

The entry $(q, q')$ of the $(Q+1) \times (Q+1)$ sensitivity matrix for the dispersion parameters is given by

$$\mathrm{S}_{\boldsymbol{\lambda}_{qq'}} = \mathrm{E}\left( \frac{\partial}{\partial \lambda_{q'}} \psi_{\boldsymbol{\lambda}_q}(\boldsymbol{\lambda},\boldsymbol{\beta}) \right) = -\sum_{i=1}^{n} \omega_{i\lambda_q} \sigma_i \omega_{i\lambda_{q'}} \sigma_i, \tag{5}$$

where $\lambda_q$ and $\lambda_{q'}$ denote both $\gamma_q$ and $p$. In a similar way, the cross entries of the sensitivity matrix are given by

$$\mathrm{S}_{\boldsymbol{\beta}_s \boldsymbol{\lambda}_q} = \mathrm{E}\left( \frac{\partial}{\partial \lambda_q} \psi_{\boldsymbol{\beta}_s}(\boldsymbol{\beta},\boldsymbol{\lambda}) \right) = 0 \tag{6}$$

and

$$S_{\lambda_q \beta_s} = E\left(\frac{\partial}{\partial \beta_s}\psi_{\lambda_q}(\boldsymbol{\lambda},\boldsymbol{\beta})\right) = -\sum_{i=1}^{n} \omega_{i\lambda_q}\sigma_i\omega_{i\beta_s}\sigma_i, \tag{7}$$

where $\omega_{i\beta_s} = -\partial\sigma_i^{-1}/\partial\beta_s$.

Finally, the joint sensitivity matrix for $\boldsymbol{\theta}$ is given by

$$S_{\boldsymbol{\theta}} = \begin{pmatrix} S_{\boldsymbol{\beta}} & \mathbf{0} \\ S_{\boldsymbol{\lambda}\boldsymbol{\beta}} & S_{\boldsymbol{\lambda}} \end{pmatrix}, \tag{8}$$

whose entries are defined in eqs. (2), (5), (6) and (7).

The asymptotic variance of the estimating function estimators denoted by $\hat{\boldsymbol{\theta}}$ are obtained by the inverse of the Godambe information matrix. The Godambe information matrix is given by $J_{\boldsymbol{\theta}} = S_{\boldsymbol{\theta}} V_{\boldsymbol{\theta}}^{-1} S_{\boldsymbol{\theta}}^{\top}$, where $\top$ denotes transpose.

The variability matrix for $\boldsymbol{\theta}$ has the form

$$V_{\boldsymbol{\theta}} = \begin{pmatrix} V_{\boldsymbol{\beta}} & V_{\boldsymbol{\beta}\boldsymbol{\lambda}} \\ V_{\boldsymbol{\lambda}\boldsymbol{\beta}} & V_{\boldsymbol{\lambda}} \end{pmatrix}, \tag{9}$$

where $V_{\boldsymbol{\lambda}\boldsymbol{\beta}} = V_{\boldsymbol{\beta}\boldsymbol{\lambda}}^{\top}$ and $V_{\boldsymbol{\lambda}}$ depend on the third and fourth moments of $Y_i$, respectively. In order to avoid such a dependence on high order moments, we adopted the empirical version of the variability matrix obtained by

$$\tilde{V}_{\lambda_{qq'}} = \sum_{i=1}^{n} \psi_{\lambda_q}(\boldsymbol{\lambda},\boldsymbol{\beta})_i \psi_{\lambda_{q'}}(\boldsymbol{\lambda},\boldsymbol{\beta})_i \quad \text{and} \quad \tilde{V}_{\lambda_q\beta_s} = \sum_{i=1}^{n} \psi_{\lambda_q}(\boldsymbol{\lambda},\boldsymbol{\beta})_i \psi_{\beta_s}(\boldsymbol{\lambda},\boldsymbol{\beta})_i. \tag{10}$$

Thus, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, J_{\boldsymbol{\theta}}^{-1}), \tag{11}$$

where $J_{\boldsymbol{\theta}}^{-1}$ is the inverse of the Godambe information matrix.

To solve the system of equations, we adopted the `chaser` algorithm

$$\begin{aligned} \boldsymbol{\beta}^{(i+1)} &= \boldsymbol{\beta}^{(i)} - S_{\boldsymbol{\beta}}^{-1}\psi_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{(i)},\boldsymbol{\lambda}^{(i)}) \\ \boldsymbol{\lambda}^{(i+1)} &= \boldsymbol{\lambda}^{(i)} - \alpha S_{\boldsymbol{\lambda}}^{-1}\psi_{\boldsymbol{\lambda}}(\boldsymbol{\beta}^{(i+1)},\boldsymbol{\lambda}^{(i)}), \end{aligned} \tag{12}$$

in that case $\alpha$ is a tuning constant used to control the step-length. The chaser algorithm uses the insensitivity property, see eq. (6), which allows us to use two separate equations to update $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. For further details, we refer the interested reader to Bonat and Jørgensen [21], Bonat [29] and Jørgensen and Knudsen [28]. The algorithm presented in this section is easily implemented in `R` through the `mcglm` package [29], whose code is available as a supplementary material.

## 5 Simulation studies

In this section we shall present two simulation studies that are conducted to verify the properties of the estimating function estimators and highlight the importance of the correct modeling of the dispersion structure.

### 5.1 Properties of the estimating function estimators

We carry out a simulation study to check the properties of the proposed estimating function estimators in a finite sample scenario. The regression model for the mean structure was defined as $\mu_i = \exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\}$, where $\boldsymbol{\beta} = (\log(10), 1.0, 0.5)^{\top}$. The covariates $x_{1i}$ and $x_{2i}$ were generated from standard Gaussian and Bernoulli($p = 0.5$) distributions, respectively. The parameter values were picked in order to have counts from 0 to approximately 350.

The focus of this paper is on the dispersion structure, consequently, we designed three simulation scenarios by considering different levels of overdispersion (low, medium and large) and combined them with different

values of the power parameter, see Table 1. For each scenario, we generated 1000 data sets considering four sample sizes (100, 250, 500 and 1000). The dispersion regression model consists of two covariates and the intercept. We decided to use the same set of covariates to model both the mean and dispersion structures.

**Table 1:** Power parameter and linear predictor for the dispersion regression model by simulation scenario.

| Power parameter | Scenario | Dispersion regression model |
|---|---|---|
| $p = 0$ | Low | $\phi_i = \exp\{\log(2.0) + 0.2x_{1i} + 0.3x_{2i}\}$ |
| | Medium | $\phi_i = \exp\{\log(2.5) + 0.5x_{1i} + 0.5x_{2i}\}$ |
| | Large | $\phi_i = \exp\{\log(4.5) + 0.8x_{1i} + 0.8x_{2i}\}$ |
| $p = 1.01$ | Low | $\phi_i = \exp\{\log(0.5) + 0.1x_{1i} - 0.8x_{2i}\}$ |
| | Medium | $\phi_i = \exp\{\log(2.0) + 0.8x_{1i} + 0.5x_{2i}\}$ |
| | Large | $\phi_i = \exp\{\log(3.5) + 1.5x_{1i} + 1.2x_{2i}\}$ |
| $p = 1.5$ | Low | $\phi_i = \exp\{\log(0.3) + 0.1x_{1i} + 0.3x_{2i}\}$ |
| | Medium | $\phi_i = \exp\{\log(1.5) + 0.5x_{1i} + 0.5x_{2i}\}$ |
| | Large | $\phi_i = \exp\{\log(2.0) + 1.2x_{1i} + 1.3x_{2i}\}$ |
| $p = 2$ | Low | $\phi_i = \exp\{\log(0.2) + 0.1x_{1i} + 0.3x_{2i}\}$ |
| | Medium | $\phi_i = \exp\{\log(1.2) + 0.5x_{1i} + 0.3x_{2i}\}$ |
| | Large | $\phi_i = \exp\{\log(1.7) + 0.8x_{1i} + 1.0x_{2i}\}$ |
| $p = 3$ | Low | $\phi_i = \exp\{\log(0.003) + 0.001x_{1i} + 0.005x_{2i}\}$ |
| | Medium | $\phi_i = \exp\{\log(0.01) + 0.005x_{1i} + 0.01x_{2i}\}$ |
| | Large | $\phi_i = \exp\{\log(0.08) + 0.01x_{1i} + 0.2x_{2i}\}$ |

Figure 3 and Figure 4 present the average bias plus and minus the average standard errors (SE) for the regression and dispersion parameters under each scenario. The scales are standardized for each parameter by dividing the average bias and the limits of the confidence intervals by the SE obtained for the sample of size 100.

The results in Figure 3 and Figure 4 show that for all simulation scenarios both the average bias and standard errors tend to 0 as the sample size is increased. These results illustrate the consistency and unbiasedness (for large sample) of the estimating function estimators of the regression and dispersion coefficients. As expected, the scenarios with larger overdispersion and higher power parameter values are the most challenging for the fitting algorithm, consequently, in these scenarios larger samples are required for the correct estimation. It is also clear that the estimation of the dispersion coefficients is harder than the regression coefficients and larger samples are required to reach unbiased estimates. Concerning the estimation of the power parameter, we note that only in the large overdispersion case and $p = 3$ scenario our fitting algorithm did not provide unbiased estimates for large samples.

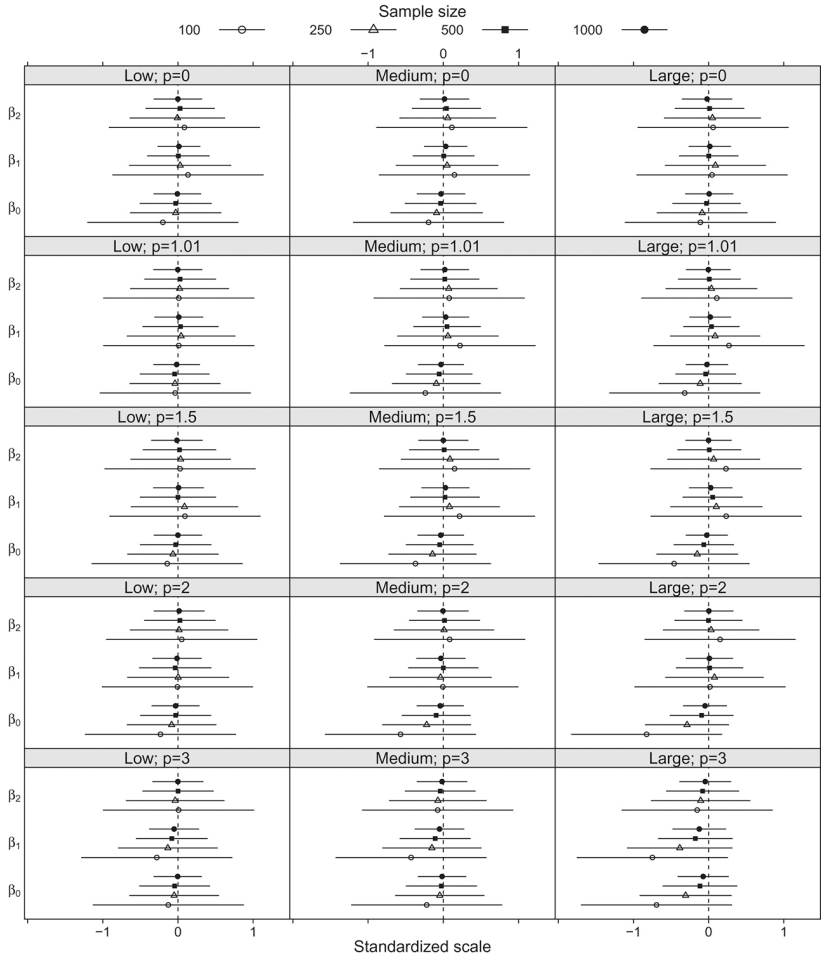


**Figure 3:** Average bias and confidence intervals on a standardized scale by sample size and simulation scenario - Regression coefficients.

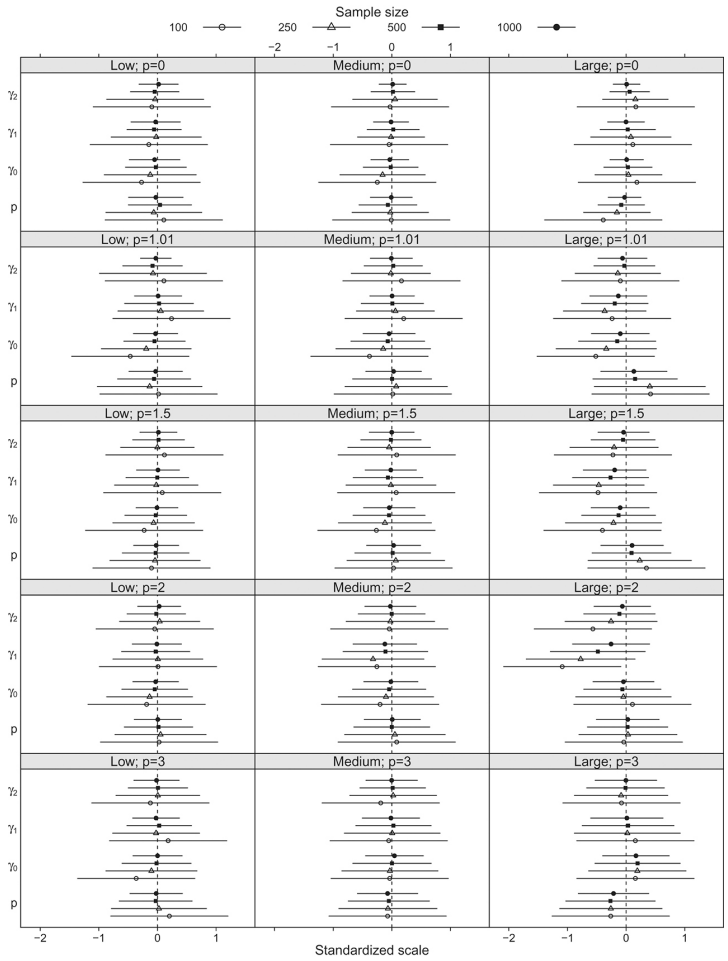


**Figure 4:** Average bias and confidence intervals on a standardized scale by sample size and simulation scenario - Dispersion and power coefficients.

In order to further investigate the properties of our estimating functions estimators Figure 5 presents the empirical coverage rate for all model parameters, sample sizes and simulation scenarios.

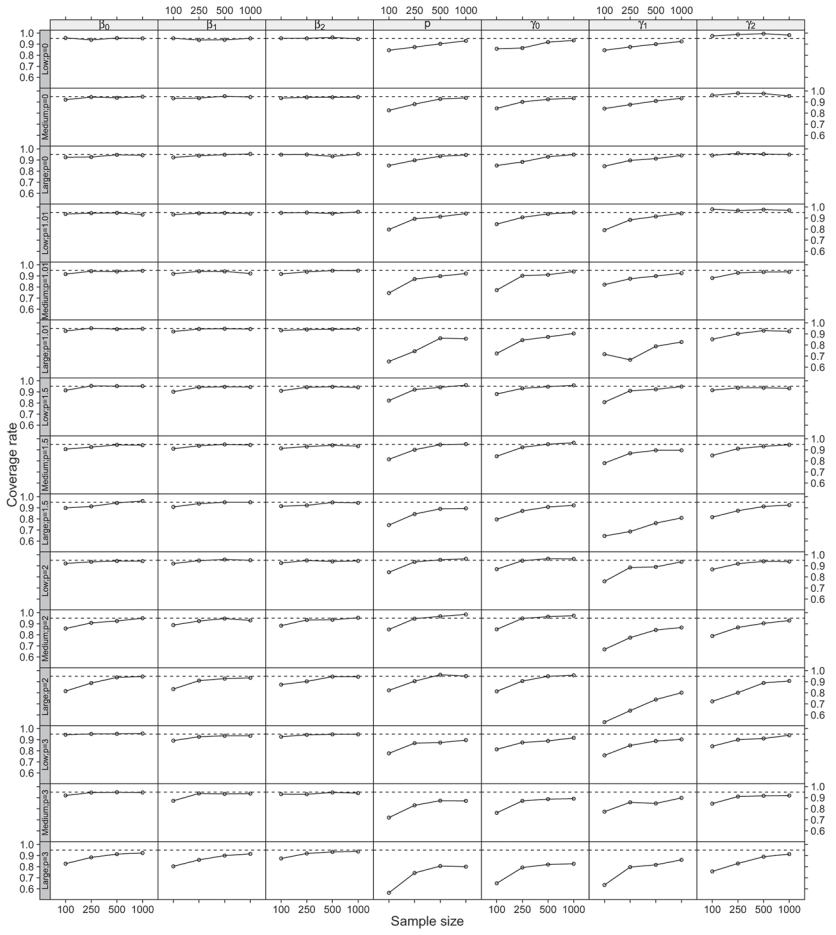


**Figure 5:** Coverage rate for each parameter by sample sizes and simulation scenarios.

The empirical coverage rates for the regression coefficients are close to the nominal level of 95% for all sample sizes and simulation scenarios. Regarding the power and dispersion parameters the empirical coverage rates present values close to the nominal levels for large samples; however, for small samples the empirical coverage rate tends to be slightly lower than the nominal level of 95%. The worst results appear for the power parameter at low overdispersion levels.

In Figure 3 and Figure 4, we opted to present the average bias and standard error in standardized scale. Although, this approach is convenient to show the properties of the estimators, it can be misleading because the size of the estimator's variance is standardized to be 1. Thus, in the supplementary material we present Figure 3 and Figure 4, however the horizontal lines (confidence intervals) were replaced by the minimum and maximum estimated values. These results agree that our estimators are consistent. Furthermore, they highlighted that for small samples the uncertainty around the dispersion coefficients are quite large, mainly for power parameter values close to 0.

## 5.2 Impact of misspecification of the dispersion structure

The goal of this simulation study is to highlight the importance of correctly modeling the dispersion structure. We simulated 1000 data sets from the double Poisson-Tweedie regression model and fitted both the double Poisson-Tweedie (correct model) and the Poisson-Tweedie regression model ignoring the covariates in the dispersion structure (incorrect model). Then, we evaluated the relative efficiency which was defined as the ratio of the standard errors obtained from the incorrect (numerator) and correct (denominator) models.

We fixed the power parameter at the values $p = 1.01$, 1.5, 2 and 3. The mean structure was specified as

$$\mu_i = \exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{32} x_{32i} + \beta_{33} x_{33i} + \beta_{34} x_{34i} + \beta_{35} x_{35i}\}, \tag{13}$$

where $\boldsymbol{\beta} = (\log(10), 1, -2, 0.8, 1.5, 0.5, -1)^\top$. Similar to the first simulation study, the regression coefficients were fixed in order to have counts approximately from 0 to 350. The covariates $x_{1i}$ and $x_{2i}$ were generated from a Gaussian (mean zero and variance $0.3^2$) and Bernoulli ($p = 0.7$) distributions, respectively. The covariates $x_{3ji}$

for $j = 2, \ldots 5$ are dummies representing a factor with five levels of equal sample sizes. Finally, the dispersion structure was specified as

$$\phi_i = \exp\{\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_{32} x_{32i} + \gamma_{33} x_{33i} + \gamma_{34} x_{34i} + \gamma_{35} x_{35i}\}, \tag{14}$$

where $\boldsymbol{\gamma} = (2.5, 2, -1.3, -0.8, -1.5, -0.5, 1)^\top$. The covariates $z_{1i}$ and $z_{2i}$ were simulated from Gaussian (mean zero and variance $0.5^2$) and Bernoulli ($p = 0.5$) distributions, respectively. It is important to highlight that the covariates $x_{3ji}$ appear in both mean and dispersion structures.

In order to evaluate the impact of the misspecification of the dispersion structure on the regression coefficients, we fitted the correct model, i.e the double Poisson-Tweedie regression model and the naive one i.e the orthodox Poisson-Tweedie regression model.

We defined the relative efficiency as the ratio of the standard errors obtained from the incorrect and correct models. Thus, values larger than one indicate that the correct model is more efficient (smaller standard errors) than the incorrect one. Figure 6 shows the relative efficiency for each regression parameter and simulation scenario.

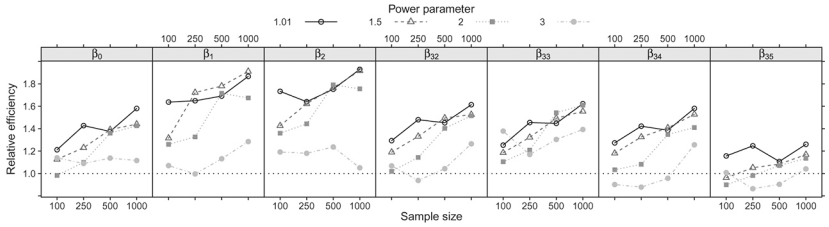


**Figure 6:** Relative efficiency for each parameter by power parameter and sample size.

Results in Figure 6 show that the correct model is more efficient than the incorrect model for all simulation scenarios and regression parameters with exception of the coefficient $\beta_{35}$ in the case of $p = 3$. Thus, it is clear that the correct specification of the dispersion structure has a strong impact on the standard errors associated to the regression coefficients. This result highlights the importance of correctly modeling the dispersion structure. In Figure 6, we used the average of the estimated standard errors obtained based on the Godambe information matrix. In the supplementary material we provide the same Figure, but with standard errors computed based on the Monte Carlo (empirical) variance. Overall the results are quite similar.

## 6 Analyzing CD4 counts in HIV-positive pregnant women

In this section, we apply the class of double Poisson-Tweedie regression models for count data to analyze the data set presented in Section 2. The second-moments assumptions of the double Poisson-Tweedie regression models require the specification of linear predictors for modeling the mean and dispersion structures. In this application, for composing the linear predictors we have three continuous covariates, namely `TART` - how long the patient has been on antiretroviral therapy (ART) in weeks; `GAART` - gestational age at which the patient started ART and `AGE` - patient age. Additionally, we have nine categorical covariates, namely `GA` - gestational age in weeks; `HIVD` - HIV diagnostic during pregnancy; `CR` - city of residence; `VL` - viral load; `CO` - co-infection; `NBC` - newborn condition; `HIVRF` - known HIV risk factors; `DEL` - delivery and `URG` - use of raltegravir (see Section 2).

We adopted a stepwise type procedure for selecting the components of the linear predictors. The Wald test was used in the forward and backward steps. Our strategy to select the final model consists of: (i) selecting the components of the mean linear predictor, (ii) selecting the components of the dispersion linear predictor and (iii) removing non-significant effects (if any) in both mean and dispersion linear predictors. We highlight that first, we selected the covariates to compose the mean model, i.e. in this step we consider the dispersion constant. Then, we fixed the selected covariates of the mean model in order to select the covariates for composing the dispersion model. Finally, we evaluated the final model and drop (if any) non-significant covariates in both mean and dispersion structures. The final mean and dispersion linear predictors are given, respectively, by

$$\log(\mu_i) = \beta_0 + \beta_1 \mathrm{VL}(\leq 999)_i + \beta_2 \mathrm{VL}(\geq 1000)_i + \beta_3 \mathrm{HIVD}_i + \beta_4 \mathrm{CR}_i$$

and

$$\log(\phi_i) = \gamma_0 + \gamma_1 \mathrm{TART}_i.$$

On the adopted parametrization, $\beta_0$ is associated with undetectable viral load and $\beta_1$ and $\beta_2$ are differences for viral load $\leq 999$ and $\geq 1000$, respectively. Similarly, $\beta_3$ is the effect of HIV diagnostics during pregnancy and $\beta_4$ measures the effect of living out of the capital Curitiba, i.e. the reference level is Curitiba. The linear predictor for the dispersion structure is composed by an intercept and the effect of the continuous covariate `TART`.

In order to further investigate the effect of the estimation of the power parameter, we opted to fit the double Poisson-Tweedie regression model as well as some of its main special cases. Note that, the covariates selection was done based on the more general model, i.e. the double Poisson-Tweedie with power parameter estimated based on data. Then, we fit the special cases obtained by fixing the power parameter at the values 0 (Hermite), 1 (Neyman Type A), 1.5 (Pólya Aeppli), 2 (negative binomial) and 3 (Poisson inverse-Gaussian) keeping the mean and dispersion structures as obtained based on the double Poisson-Tweedie regression model. Table 2 presents the pseudo version of the *Akaike* (`pAIC`) and Bayesian (`pBIC`) information criterion along with the maximized value of the Gaussian pseudo log-likelihood (`plogLik`) and the number of parameters (`np`) involved in the fit [29].

**Table 2:** Pseudo *Akaike* (`pAIC`) and Bayesian (`pBIC`) information criterion along with maximized Gaussian pseudo log-likelihood (`plogLik`) and number of parameters (`np`) for alternative models.

| Models | pAIC | pBIC | plogLik | np |
|---|---|---|---|---|
| Double Poisson-Tweedie | 3878.44 | 3907.54 | −1931.22 | 8 |
| Hermite ($p = 0$) | 3876.58 | 3902.04 | −1931.29 | 7 |
| Neyman Type A ($p = 1$) | 3880.40 | 3905.87 | −1933.20 | 7 |
| Pólya Aeppli ($p = 1.5$) | 3887.26 | 3912.72 | −1936.63 | 7 |
| Negative binomial ($p = 2$) | 3898.26 | 3923.76 | −1942.15 | 7 |
| Poisson inverse-Gaussian ($p = 3$) | 3936.58 | 3962.04 | −1961.29 | 7 |

Results in Table 2 show that the fit of the double Poisson-Tweedie regression model $\hat{p} = 0.18(0.47)$ is quite similar to the Hermite and Neyman Type A models in terms of `plogLik`. However, since the double Poisson-Tweedie has the extra power parameter, the `pAIC` and `pBIC` criterion indicate the Hermite as the best fit. Furthermore, if we consider the `pBIC` criterion the Hermite and Neyman Type A models provide better fit than the proposed double Poisson-Tweedie regression. However, the presented measures of goodness-of-fit are quite limited, because they do not offer a measure of uncertainty. For example, it is challenging to decide which model Hermite or Neyman Type A fits better to the data, because the difference in terms of pseudo log-likelihood values is quite small. To better explore these results, Table 3 presents the corresponding estimates and standard errors obtained by the fit of the Hermite, Neyman Type A, Pólya Aeppli, negative binomial and Poisson inverse-Gaussian models.

**Table 3:** Parameter estimates and standard errors for the double Poisson Tweedie (DPT), Hermite (HMT), Neyman type A (NTA), Pólya Aeppli (PA), negative binomial (NB) and Poisson inverse-Gaussian (PIG) models.

| Parameter | DPT | HMT | NTA | PA | NB | PIG |
|---|---|---|---|---|---|---|
| $\beta_0$ | 6.40(0.05)* | 6.39(0.05)* | 6.39(0.05)* | 6.39(0.05)* | 6.40(0.05)* | 6.40(0, .05)* |
| $\beta_1$ | −0.23(0.06)* | −0.23(0.06)* | −0.23(0.06)* | −0.24(0.06)* | −0.24(0.06)* | −0.25(0.07)* |
| $\beta_2$ | −0.40(0.08)* | −0.40(0.08)* | −0.42(0.08)* | −0.43(0.07)* | −0.44(0.07)* | −0.46(0.07)* |
| $\beta_3$ | 0.15(0.05)* | 0.15(0.05)* | 0.16(0.05)* | 0.17(0.05)* | 0.17(0.05)* | 0.18(0.06)* |
| $\beta_4$ | −0.14(0.05)* | −0.14(0.05)* | −0.15(0.06)* | −0.16(0.06)* | −0.17(0.06)* | −0.19(0.06)* |
| $\gamma_0$ | 10.10(2.98)* | 11.27(0.15)* | 4.82(0.15)* | 1.60(0.15)* | −1.60(0.15)* | −7.99(0.18)* |
| $\gamma_1$ | −0.02(0.01)* | −0.02(0.01)* | −0.01(0.01)* | −0.01(0.01) | −0.00(0.01) | 0.01(0.01) |
| $p$ | 0.18(0.47) | 0 | 1 | 1.5 | 2 | 3 |

Table 3 shows that in general all models agree in terms of the significance of the regression coefficients, however, only the double Poisson-Tweedie, Hermite and Neyman Type A detected the significance of the covariate `TART` on the dispersion regression model. This result agrees with the goodness-of-fit measures presented in Table 2 where we noted that the fits of these three models are really similar and better than the Pólya Aeppli, negative binomial and Poisson inverse-Gaussian fits.

It is easy to explain such a similarity by analyzing the power parameter estimate. Recall, that the power parameter is an index that distinguishes between the special cases of the Poisson-Tweedie family of distributions.

In this data analysis, the estimated value was $\hat{p} = 0.18$, which is close to 0 explaining the similarity of the double Poisson-Tweedie and Hermite models. However, the standard error associated with this estimate was 0.48, thus the associated 95% confidence interval is given by $(-0.74|1.10)$, which shows that both Hermite and Neyman Type A models provide a reasonable fit to the data. These results show that the double Poisson-Tweedie model provides the best fit as well as automatically adapted to the underlying count distribution, without the need of goodness-of-fit measures as presented in Table 2. Consequently, the double Poisson-Tweedie model is robust against model misspecification.

To further investigate the effect of model misspecification, Figure 7 presents the estimates (A) and the relative standard errors (B) obtained by fitting the double Poisson-Tweedie regression model with different fixed power parameter values. We highlight that the double Poisson-Tweedie regression model with the power parameter estimated based on the data is used as the reference for the computation of the relative standard errors.

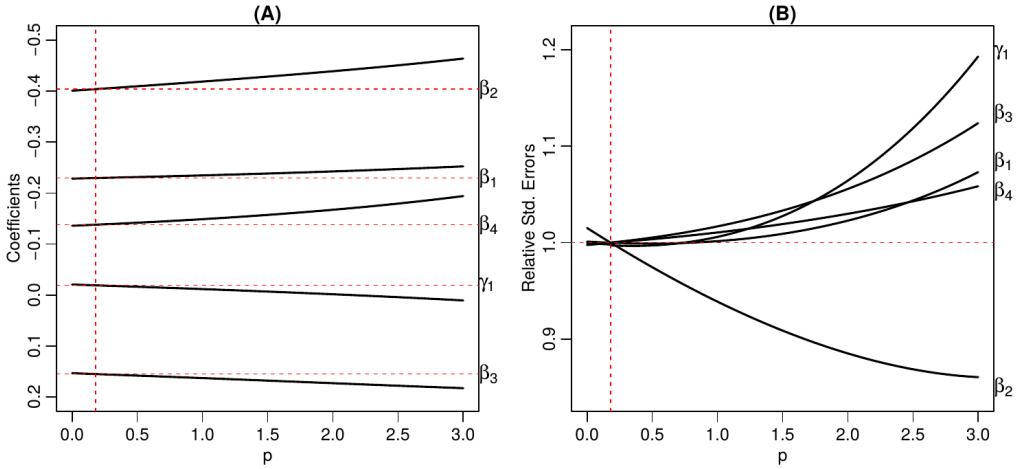


**Figure 7:** Illustration of the change in the regression and dispersion parameter estimates for different values of $p$ (A). Illustration of the change in the relative standard errors of the regression and dispersion parameter estimates for different values of $p$ (B).

Figure 7 shows that in general the regression and dispersion parameter estimates are only barely affected by the power parameter. On the other hand, the associated standard errors are strongly affected by the power parameter, with changes reaching up to 20%. Figure 7 should be interpreted such that if $p = 3$ were used, the estimated standard error of $\gamma_1$ would be approximately 20% larger than the value obtained from the best fit. It highlights the importance of a joint estimation of the regression, dispersion and power parameters when fitting double Poisson-Tweedie regression models.

## 7 Discussion

In this paper, we have first presented the double Poisson-Tweedie regression models for analyzing count data. The models are based only on second-moments assumptions and allow to model the mean and dispersion structures in a regression model fashion. Estimation and inference is easily done using an estimating function approach in the style of Wedderburn's quasi-likelihood [22]. Simulation studies showed that the proposed estimating function approach provides consistent estimators for both mean and dispersion coefficients and highlighted the importance of correctly modeling the dispersion structure. Furthermore, we showed through a data analysis that our model is quite flexible and can easily adapt to the underlying count distribution without the need of additional goodness-of-fit measures. It is important to highlight that the methodology presented in this article can be used as a test for the assumption of dispersion homogeneity.

An advantage of the estimating function approach as used in this paper combining the quasi-score and Pearson estimating functions is the insensitivity property which is an analogue to the orthogonality property in the context of maximum likelihood estimation. The insensitivity property allows us to apply the chaser algorithm using two separate equations to update the regression and dispersion parameters. Consequently, it simplifies the fitting process.

We have then analyzed a data set concerning CD4 counts in HIV-positive pregnant women assisted in a public hospital in Curitiba, PR, Brazil. The main goal of the data analysis was to detect factors influencing the CD4 counts. We considered a set of three continuous and nine categorical covariates. The results showed that those pregnant non-resident in the capital Curitiba, with viral load $\geq 1000$ and with previous diagnostic of HIV are the ones with lower CD4 count levels.

The city of residence is a proxy for social and economic living conditions as well as for health service access. In general, people living out of the capital have limited access to the health service network, which could explain the negative effect in the CD4 counts. Concerning the negative association between CD4 and viral load, such an association was already reported in [30]. Patients on ART keeping suppressed viral loads tend to recover their CD4 levels better than those off medication. In Brazil, all pregnant women are routinely tested for HIV during first trimester of pregnancy. This strategy provides earlier diagnosis, in the asymptomatic stage of HIV infection, usually before decreasing their CD4 cell count. It could explain that pregnant women with previous diagnostic of HIV infection present lower levels of CD4 cell count.

Finally, we have detected that the covariate `TART` which measures how long the patient has been on ART in weeks affects the dispersion structure being that longer in ART implies less dispersion. Such results agree with the fact that ART is effective on controlling the CD4 counts in HIV-positive pregnant. However, we face such result with caution, since our simulations studies showed that the standard errors for the dispersion components could be underestimated for small sample sizes as in our data analysis. Thus, further investigations are required for a definitive conclusion.

Some possible topics for future research include the extension of the double Poisson-Tweedie regression models to deal with non-independent data in the Liang and Zeger style [31] as well as to further study the model properties to deal with multiple count response variable as proposed in [21, 32]. There is also a need to develop methods for model checking such as residual analysis, leverage and outliers detection.

## References


[1] WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Technical report. World Health Organization, 2016.

[2] UNAIDS, J U. N. P. oH. On the Fast-Track to end AIDS by 2030: Focus on location and population. Technical report. Joint United Nations Programme on HIV/AIDS, 2015.

[3] Landefeld CC, Fomenou LA, Ateba F, Msellati P. Prevention of mother-to-child transmission of HIV in Yaounde: Barrier to care. AIDS care. 2018;30:116–20.

[4] French CE, Thorne C, Byrne L, Cortina-Borja M, Tookey PA. Presentation for care and antenatal management of HIV in the UK 2009–2014. HIV Med. 2017;18:161–70.

[5] Grover G, Vajala R, Swain PK. On the assessment of various factors effecting the improvement in CD4 count of aids patients undergoing antiretroviral therapy using generalized poisson regression. J Appl Stat. 2015;42:1291–305.

[6] Lok JJ, Bosch RJ, Benson CA, Collier AC, Robbins GK, Shafer RW, et al. Long-term increase in CD4+ T-cell counts during combination antiretroviral therapy for HIV-1 infection. AIDS (London, England). 2010;24:1867–76.

[7] Seyoum A, Zewotir T. Quasi-Poisson versus negative binomial regression models in identifying factors affecting initial CD4 cell count change due to antiretroviral therapy administered to HIV-positive adults in North–West Ethiopia. AIDS Res Ther. 2016;13:2–10.

[8] Helleberg M, Kronborg G, Larsen CS, Pedersen G, Pedersen C, Obel N, et al. CD4 decline is associated with increased risk of cardiovascular disease, cancer, and death in virally suppressed patients with HIV. Clin Infec Dis. 2013;57:314–21.

[9] Cameron AC, Trivedi PK.. Regression analysis of count data, vol. 53 Cambridge: Cambridge University Press, 2013

[10] Zeviani WM, Ribeiro Jr. PJ, Bonat WH, Shimakura SE, Muniz JA. The Gamma-count distribution in the analysis of experimental underdispersed data. J Appl Stat. 2014;41:2616–26.

[11] Bonat WH, Jørgensen B, Kokonendji CC, Hinde J, Demétrio CG. Extended Poisson-Tweedie: properties and regression models for count data. Stat Modell. 2018;18:24–49.

[12] El-Shaarawi AH, Zhu R, Joe H. Modelling species abundance using the Poisson-Tweedie family. Environmetrics. 2011;22:152–64.

[13] Hinde J, Demétrio CG. Overdispersion: models and estimation. Comput Stat Data Anal. 1998;27:151–70.

[14] Kokonendji CC, Demétrio CG, Zocchi SS. On Hinde–Demétrio regression models for overdispersed count data. Stat Method. 2007;4:277–91.

[15] Mahmoodi M, Moghimbeigi A, Mohammad K, Faradmal J. Semiparametric models for multilevel overdispersed count data with extra zeros. Stat Method Med Res. 2016;27:1187–201.

[16] Oliveira M, Einbeck J, Higueras M, Ainsbury E, Puig P, Rothkamm K. Zero-inflated regression models for radiation-induced chromosome aberration data: a comparative study. Biometric J. 2016;58:259–79.

[17] Rigby RA, Stasinopoulos DM, Akantziliotou C. A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. Comput Stat Data Anal. 2008;53:381–93.

[18] Sellers KF, Shmueli G. A flexible regression model for count data. Annals Appl Stat. 2010;4:943–61.

[19] Smyth GK. Generalized linear models with varying dispersion. J R Stat Soc Ser B Method. 1989;51:47–60.

[20] Andersen DA, Bonat WH. Double generalized linear compound Poisson models to insurance claims data. Electron J Appl Stat Anal. 2017;10:384–407.

[21] Bonat WH, Jørgensen B. Multivariate covariance generalized linear models. J R Stat Soc: Ser C (Appl Stat). 2016;65:649–75.

[22] Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the gauss–newton method. Biometrika. 1974;61:439–47.

[23] Esnaola M, Puig P, Gonzalez D, Castelo R, Gonzalez JR. A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated rna-seq experiments. BMC Bioinf. 2013;14:254–76.

[24] Kokonendji CC, Dossou-Gbété S, Demétrio CG. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. Stat Oper Res Trans. 2004;28:201–14.
[25] Moria D, Higueras M, Puig P, Oliveira M. hermite: generalized Hermite distribution. https://CRAN.R-project.org/package=hermite, r package version 1.1.1. 2015.
[26] Jørgensen B, Kokonendji CC. Discrete dispersion models and their tweedie asymptotics. AStA Adv Stat Anal. 2016;100:43–78.
[27] Cox DR, Hinkley DV. Theoretical statistics. London, England: Chapman & Hall, 1974.
[28] Jørgensen B, Knudsen SJ. Parameter orthogonality and bias adjustment for estimating functions. Scand J Stat. 2004;31:93–114.
[29] Bonat WH. Multiple response regression models in R: the mcglm package. J Stat Software. 2018;85:1–30.
[30] Yu T, Wu L. Robust modelling of the relationship between CD4 and viral load for complex AIDS data. J Appl Stat. 2018;45:367–83.
[31] Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73:13–22.
[32] Bonat WH, Olivero J, Grande-Vega M, Farfán MA, Fa JE. Modelling the covariance structure in marginal multivariate count models: hunting in bioko island. J Agr Biol Environ Stat. 2017;22:446–64.