Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2018)

Discrete Weibull generalized additive model: an application to count fertility data

Alina Peluso, Veronica Vinciotti and Keming Yu

Brunel University London, Uxbridge, UK

[Received January 2018. Revised August 2018]

Summary. Fertility plans, measured by the number of planned children, have been found to be affected by education and family background via complex tail dependences. This challenge was previously met with the use of non-parametric jittering approaches. The paper shows how a novel generalized additive model based on a discrete Weibull distribution provides partial effects of the covariates on fertility plans which are comparable with jittering, without the inherent drawback of conditional quantiles crossing. The model has some additional desirable features: both overdispersed and underdispersed data can be modelled by this distribution, the conditional quantiles have a simple analytic form and the likelihood is the same as that of a continuous Weibull distribution with interval-censored data. Because the likelihood is like that of a continuous Weibull distribution, efficient implementations are already available, in the R package gamlss, for a range of models and inferential procedures, and at a fraction of the time compared with the jittering and Conway–Maxwell–Poisson approaches, showing potential for the wide applicability of this approach to the modelling of count data.

Keywords: Count data; Discrete Weibull distribution; Generalized additive model; Planned fertility

1. Introduction

Fertility plans measured by the number of planned children, or ideal fertility, have been previously found to be affected by education and family background (Knodel and Prachuabmoh, 1973; Pritchett, 1994). In a recent study in Mexico, Miranda (2008) showed how the dependence of ideal fertility on education and family background is complex, with effects mostly at the tail of the distribution, and how ideal fertility is typically underdispersed relative to a Poisson distribution. In this paper, using the latest data from the Mexican national survey of demographic dynamics conducted by the Instituto Nacional de Estadística y Geografía (Instituto Nacional de Estadística y Geografía, 2014), we propose a novel regression model to discover determinants of planned fertility and to quantify this dependence.

Methods to address questions such as this fall in the general area of regression analysis of count data, with many applications ranging from healthcare, biology, social science, marketing and crime data analyses (Cameron and Trivedi, 2013; Hilbe, 2014). Among these methods, generalized linear models (Nelder and Wedderburn, 1972) are popular in the parametric literature. Here, the conditional distribution of the response variable given the predictors is assumed to follow a specified distribution, with the conditional mean linked to the predictors via a regression model. For example, Poisson regression assumes that the conditional distribution is Poisson with a conditional mean regressed on the covariates through the log-link function. Although

Address for correspondence: Veronica Vinciotti, Department of Mathematics, Brunel University London, Uxbridge, UB8 3PH, UK.
E-mail: veronica.vinciotti@brunel.ac.uk

Poisson regression is fundamental to the regression analysis of count data, it is often of limited use for real data, because of its property of equal mean and variance. Real data usually present overdispersion relative to a Poisson distribution or the opposite case of underdispersion. Negative binomial regression is widely considered as the default choice for data that are overdispersed relative to a Poisson distribution, although other options, such as the Poisson–inverse Gaussian model (Willmot, 1987), are available. However, neither negative binomial regression nor the Poisson–inverse Gaussian model can deal with data that are underdispersed relative to a Poisson distribution. These can arise in various applications, such as in cases where the data are preprocessed for confidentiality (Kadane *et al.*, 2006). There have been some attempts to extend Poisson-based models to include also underdispersion, such as the generalized Poisson regression model (Consul and Famoye, 1992), Conway–Maxwell–Poisson regression (Sellers and Shmueli, 2010), extended Poisson process models (Smith and Faddy, 2016) or hyper-Poisson regression models (Sáez-Castillo and Conde-Sánchez, 2013). These models are all modifications of a Poisson model and have been shown to be quite complex and computationally intensive in practice (Chanialidis *et al.*, 2018).

At the other end of the spectrum of parametric approaches, quantile regression approaches focus on modelling individual quantiles of the distribution and linking these to the predictors via a regression model. Of particular note for discrete responses are the quantile regression models for binary and multinomial responses of Manski (1985) and Horowitz (1992), and the median regression approach with ordered response of Lee (1992). For a general discrete response, the literature on quantile regression for counts is mainly dominated by the jittering approach of Machado and Santos Silva (2005), which was also rephrased in a Bayesian framework by Lee and Neocleous (2010) in the context of an environmental epidemiology study. In these approaches, the fitted regression parameters are specific to the selected conditional quantile, thanks to the use of quantile-specific loss functions. Performing inference across a range of quantiles provides a global picture of the conditional distribution of the response variable, without having to specify the parametric form of the conditional distribution. This has proven to be quite useful in practice, particularly in cases where the relationship between the response and predictors is complex. This was in fact found in the planned fertility data set of Miranda (2008), whereby a jittering approach revealed effects mostly at the tails of the conditional distribution.

Quantile regression approaches, however, suffer from some drawbacks: inference must be made for each individual quantile, separate quantiles may cross and, in the case of jittering, the underlying uniform random sampling can generate instability in the estimation. The parametric literature, however, has addressed more complex dependences either by developing new distributions with additional parameters, e.g. the generalized gamma approach of Noufaily and Jones (2013) for continuous responses, or by adopting more flexible non-parametric functions to link the parameters of the conditional distribution to the covariates, leading to semiparametric approaches. Among this second line of research, one of the most popular approaches is the generalized additive model for location, scale and shape (Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2017). This paper fits within this literature. In particular, we introduce a generalized additive discrete Weibull (DW) regression model. The DW distribution itself was originally developed by Nakagawa and Osaki (1975) as a discretized form of the continuous Weibull distribution, which is popular in survival analysis and failure time studies. Since then, aside from some early work on parameter estimation (Khan *et al.*, 1989; Kulasekera, 1994), and some limited use in applied contexts (Englehardt and Li, 2011; Englehardt *et al.*, 2012), there are not many other contributions in the literature. Recently, we have introduced this distribution in a simple linear regression context (Klakattawi *et al.*, 2018; Haselimashhadi *et al.*, 2018), showing several desirable features: it can model both overdispersed and underdispersed data, without being restricted to either of the two, and the conditional quantiles have a simple analytical form. Moreover, in this paper we show how the likelihood from a DW model is the same as that of a continuous Weibull distribution with interval-censored data, so efficient implementations of more complex models, such as non-linear models, mixed models and mixture models, are already available in the R package `gamlss` (Rigby and Stasinopoulos, 2005).

We present the DW distribution in Section 2 and the novel generalized additive model in Section 3. In Section 4, we assess the performance of the proposed model on a simulation study, against that of the jittering approach and existing parametric approaches. Finally, in Section 5, we show how the DW generalized additive model selected on the real data returns partial effects of planned fertility similar to those of the jittering approach, without the inherent drawback of quantile crossing, and is comparable with the fitting of a Conway–Maxwell–Poisson model, at a fraction of the time.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
https://rss.onlinelibrary.wiley.com/hub/journal/14679876/series-
c-datasets
```

2. Discrete Weibull distribution

In this section, we report some important results on the DW distribution which will be used later in the paper.

2.1. *The distribution*

If Y follows a (type 1) DW distribution (Nakagawa and Osaki, 1975), then the cumulative distribution function of Y is given by

$$F(y;q,\beta)=\begin{cases} 1-q^{(y+1)^\beta} & \text{for } y=0,1,2,3,\ldots, \\ 0 & \text{otherwise} \end{cases}$$

and its probability mass function by

$$f(y;q,\beta)=\begin{cases} q^{y^\beta}-q^{(y+1)^\beta} & \text{for } y=0,1,2,3,\ldots, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

with the parameters $0<q<1$ and $\beta>0$. Since $f(0)=1-q$, the parameter q is directly related to the percentage of 0s.

2.2. *Moments and quantiles*

It can be shown that for a DW distribution

$$E(Y)=\sum_{y=1}^{\infty} q^{y^\beta}, \tag{2}$$

$$E(Y^2)=\sum_{y=1}^{\infty}(2y-1)q^{y^\beta}=2\sum_{y=1}^{\infty} yq^{y^\beta}-E(Y),$$

for which there are no closed form expressions, but numerical approximations can be obtained on a truncated support (Barbiero, 2015).

As for quantiles, the τ-quantile of a DW distribution is given by the smallest integer $\mu_{(\tau)}$ for which $P(Y \leqslant \mu_{(\tau)}) = 1 - q^{(y+1)^\beta} \geqslant \tau$. This gives

$$\mu_{(\tau)} = \lceil \mu_{(*\tau)} \rceil = \left\lceil \left\{ \frac{\log(1-\tau)}{\log(q)} \right\}^{1/\beta} - 1 \right\rceil, \tag{3}$$

with $\lceil \cdot \rceil$ the ceiling function. From this

$$\log(\mu_{(*\tau)} + 1) = \frac{1}{\beta} \log\{-\log(1-\tau)\} - \frac{1}{\beta} \log\{-\log(q)\}. \tag{4}$$

Given that Y is non-negative and that the cumulative density function is $1-q$ at 0, the quantile is defined only for $\tau \geqslant 1-q$. As a special case, the median of a DW distribution is given by

$$\mu_{(0.5)} = \left\lceil \left\{ -\frac{\ln(2)}{\ln(q)} \right\}^{1/\beta} - 1 \right\rceil. \tag{5}$$

Thus the quantiles of a DW distribution are given by simple analytical formulae.

2.3. *Likelihood and link with continuous Weibull distribution*

There is a natural link between the DW distribution and the continuous Weibull distribution with interval-censored data. The DW distribution was in fact developed as a discretized form of the continuous Weibull distribution (Chakraborty, 2015). In particular, let Y be a random variable distributed as a continuous Weibull distribution, with probability density function and cumulative density function given by

$$\begin{aligned} f_{\mathrm{W}}(y; q, \beta) &= \beta \log(q) y^{(\beta-1)} \exp\{-y^\beta \log(q)\} \qquad\qquad y \geqslant 0, \\ F_{\mathrm{W}}(y; q, \beta) &= 1 - \exp\{-y^\beta \log(q)\} \end{aligned}$$

respectively. Then one can show that

$$f(y) = F_{\mathrm{W}}(y+1) - F_{\mathrm{W}}(y) \qquad\qquad y = 0, 1, \ldots$$

where $f(y)$ is the DW probability mass function of equation (1). From this

$$\int_y^{y+1} f_{\mathrm{W}}(t) \mathrm{d}t = f(y).$$

Thus the likelihood of a continuous Weibull distribution with interval-censored data is equal to that of a DW distribution, i.e.

$$\prod_{i=1}^n f(y_i) = \prod_{i=1}^n \{F_{\mathrm{W}}(y_i + 1) - F_{\mathrm{W}}(y_i)\}.$$

2.4. *Discrete Weibull distribution accounts for overdispersion and underdispersion*

Dispersion in count data is formally defined in relation to a specified model being fitted to the data (Cameron and Trivedi, 2013). In particular, let

$$\mathrm{VR} = \frac{\text{observed variance}}{\text{theoretical variance}}.$$

So VR is the ratio between the observed variance from the data and the theoretical variance from the model. Then the data are said to be overdispersed, equidispersed or underdispersed relative to the fitted model if the observed variance is larger than, equal to or smaller than the theoretical

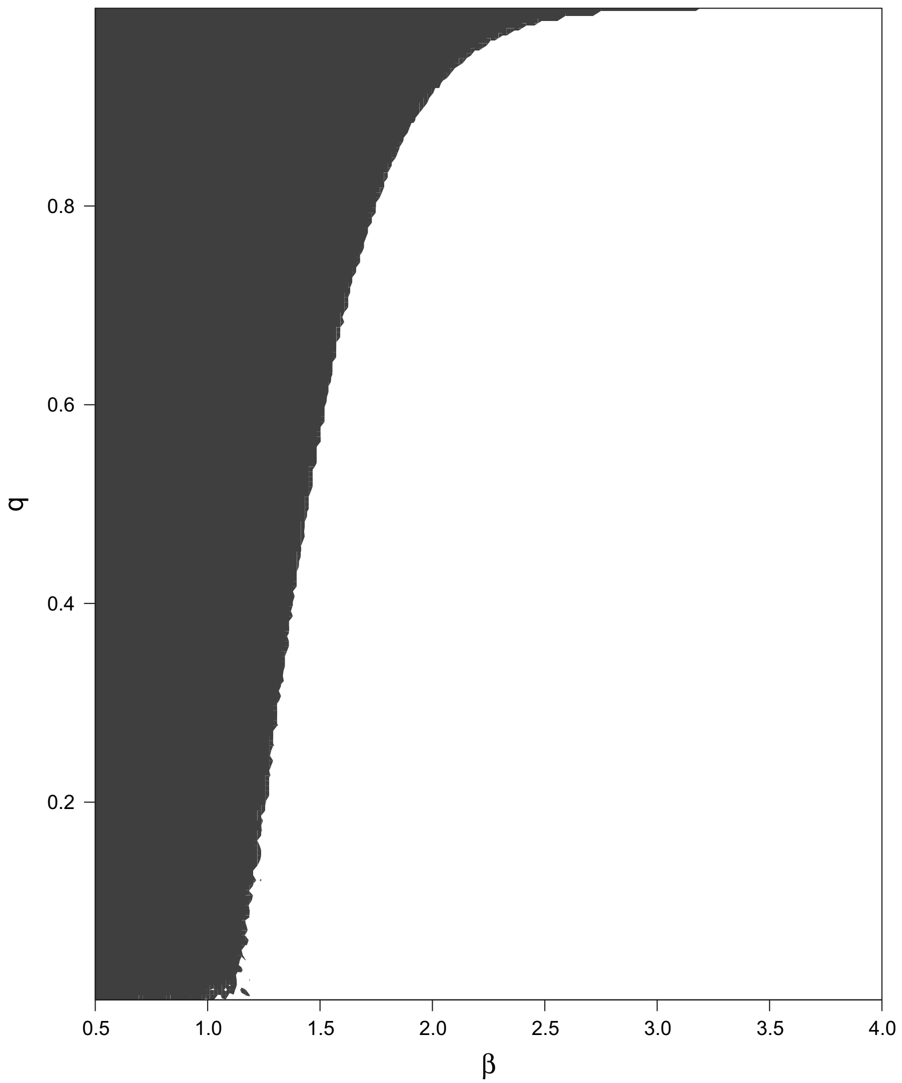


Fig. 1. Ratio of observed and theoretical variance from a Poisson model, calculated from data simulated by a DW(q, β) distribution

variance specified by the model respectively. It is common to refer to dispersion relative to a Poisson distribution. In that case, the variance of the model is estimated by the sample mean. Thus, overdispersion, equidispersion or underdispersion relative to a Poisson distribution refer to cases where the sample variance is larger than, equal to or smaller than the sample mean respectively. Since the theoretical variance of a negative binomial distribution is always greater than its mean, the negative binomial regression model is the natural choice for data that are overdispersed relative to a Poisson distribution. However, crucially, a negative binomial model cannot handle underdispersed data.

In contrast with this, Klakattawi *et al.* (2018) show how a DW distribution can handle data that are both overdispersed and underdispersed relative to a Poisson distribution. In particular, Fig. 1 shows how the DW distribution can capture cases of underdispersion, equidispersion and overdispersion relative to a Poisson distribution. Specifically, the white area corresponds to values of dispersion less than 1, i.e. underdispersed relative to a Poisson distribution, whereas the black area corresponds to overdispersion. Moreover, the plot shows that

(a) $0 < \beta \leqslant 1$ is a case of overdispersion, regardless of the value of q,
(b) $\beta \geqslant 3$ is a case of underdispersion, regardless of the value of q; in fact, the DW distribution approaches the Bernoulli distribution with mean p and variance $p(1-p)$ for $\beta \to \infty$, and
(c) $1 < \beta < 3$ leads to both cases of overdispersion and underdispersion depending on the value of q.

3. Discrete Weibull generalized additive model

3.1. Generalized additive model for location, scale and shape formulation

To capture complex dependences between the response and the covariates, such as those that we expect in our real application on planned fertility, we propose generalized additive models to link both parameters of the distribution to the covariates. In particular, we assume that the response Y has a conditional DW distribution, with the parameters q and β linked to the covariates x by

$$\log[-\log\{q(x)\}] = \sum_{p=1}^{P} f_p(x_p),$$
$$\log\{\beta(x)\} = \sum_{p=1}^{P} g_p(x_p),$$

where $x = (x_1, \ldots, x_P)$ is the vector of covariates and the functions f_p and g_p can be specified parametrically or non-parametrically. Using a basis spline representation (De Boor, 1972), the model can be written as

$$\begin{aligned}\log[-\log\{q(x)\}] &= \sum_{p=1}^{P}\sum_{d=0}^{D_p} \theta_{0pd} x_p^d + \sum_{p=1}^{P}\sum_{k=1}^{K_p} \theta_{pk}(x_p - g_{pk})^{D_p} I(x_p > g_{pk}),\\ \log\{\beta(x)\} &= \sum_{p=1}^{P}\sum_{d=0}^{D'_p} \vartheta_{0pd} x_p^d + \sum_{p=1}^{P}\sum_{k=1}^{K'_p} \vartheta_{pk}(x_p - g_{pk})^{D'_p} I(x_p > g_{pk}),\end{aligned} \tag{6}$$

where D and D' denote the polynomial degrees, K and K' are the number of knots for each covariate X_p, with g_{pk} the corresponding knots, $I(\cdot)$ is the indicator function and $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is the vector of parameters to be estimated. The log–log-link in q is motivated by the analytical formula for the quantile (equation (4)), which facilitates the interpretation of the parameters as discussed in the next subsection. Other link functions are possible, such as the logit link on q, as explored by Haselimashhadi *et al.* (2018) for the simple regression case.

Thanks to the link with the continuous Weibull likelihood that was described in Section 2, inference for these models is available in the R package `gamlss` under the `WEIic` family (Stasinopoulos and Rigby, 2007). Various model formulations and inference procedures are available in that package, such as the function `bs` for unpenalized basis splines, `cs` for smoothing cubic splines, where the coefficients are estimated under a penalty on the functions' second derivatives, and `pb` for penalized basis splines, where the differences in the coefficients of the basis functions are penalized to guarantee sufficient smoothness (Eilers and Marx, 1996).

Adding a link to both parameters means that conditional quantiles of various shapes and complexity can be captured. Considering one covariate x only, and dropping the indices p of the model for simplicity, we look closely at four cases to inspect the level of flexibility of a DW model in approximating conditional distributions.

(a) *DW linear regression model on $q(x)$ with β constant*: this model is specified as in equation (6) with $D=1$, $D'=0$ and no knots, i.e.

$$\log[-\log\{q(x)\}]=\theta_{00}+\theta_{01}x,$$
$$\log(\beta)=\vartheta_{00}.$$

Fig. 2(a) shows the case $\theta_{00}=-5$, $\theta_{01}=-3$ and $\vartheta_{00}=-1.5$, and plots $\log(\mu_{(*\tau)}+1)$ from equation (4). As expected by that equation, a linear model with β constant returns log-quantiles which are linear and parallel.

(b) *DW linear regression model on $q(x)$ and $\beta(x)$*: this model is specified as in equation (6) with $D=D'=1$ and no knots, e.g.

$$\log[-\log\{q(x)\}]=\theta_{00}+\theta_{01}x,$$
$$\log\{\beta(x)\}=\vartheta_{00}+\vartheta_{01}x,$$

for the case of a linear model on both $q(x)$ and $\beta(x)$. Fig. 2(b) shows the case $\theta_{00}=-5$, $\theta_{01}=-3$, $\vartheta_{00}=-1.5$ and $\vartheta_{01}=2$, and shows how a non-constant β enables us to obtain log-quantiles that are not parallel.

(c) *DW non-linear model for $q(x)$ with β constant*: setting $D=K=3$ and $D'=K'=0$ in equation (6) leads to a B-spline model for $q(x)$ with three interior knots:

$$\begin{aligned}\log[-\log\{q(x)\}]&=\theta_{00}+\theta_{01}x+\theta_{02}x^2+\theta_{03}x^3+\theta_1(x-g_1)^3I(x>g_1)\\&\quad+\theta_2(x-g_2)^3I(x>g_2)+\theta_3(x-g_3)^3I(x>g_3),\\ \log(\beta)&=\vartheta_{00}.\end{aligned}$$

Cubic splines are typically sufficiently complex for most real applications (Dierckx, 1995). Fig. 2(c) shows the quantiles for the cubic spline model with $\theta_{00}=-5$, $\theta_{01}=-5$, $\theta_{02}=-6$, $\theta_{03}=-4$, $\theta_1=-8$, $\theta_2=-9$, $\theta_3=-8$ and $\vartheta_{00}=-1$. The cubic spline, together with the assumption of a constant β, leads to parallel and non-linear log-quantiles, as expected by equation (4).

(d) *DW non-linear model for $q(x)$ and $\beta(x)$*: setting $D=K=D'=K'=3$ in equation (6) leads to a B-spline model for $q(x)$ and $\beta(x)$ with three interior knots:

$$\begin{aligned}\log[-\log\{q(x)\}]&=\theta_{00}+\theta_{01}x+\theta_{02}x^2+\theta_{03}x^3+\theta_1(x-g_1)^3I(x>g_1)\\&\quad+\theta_2(x-g_2)^3I(x>g_2)+\theta_3(x-g_3)^3I(x>g_3),\\ \log\{\beta(x)\}&=\vartheta_{00}+\theta_{01}x+\vartheta_{02}x^2+\vartheta_{03}x^3+\vartheta_1(x-g_1)^3I(x>g_1)\\&\quad+\vartheta_2(x-g_2)^3I(x>g_2)+\vartheta_3(x-g_3)^3I(x>g_3).\end{aligned}$$

Fig. 2(d) shows the quantiles for the cubic spline model with $\theta_{00}=-5$, $\theta_{01}=-5$, $\theta_{02}=-6$, $\theta_{03}=-4$, $\theta_1=-8$, $\theta_2=-9$ and $\theta_3=-8$, and $\vartheta_{00}=1$, $\vartheta_{01}=-1.1$, $\vartheta_{02}=-1.2$, $\vartheta_{03}=-0.5$, $\vartheta_1=-1.3$, $\vartheta_2=-1$ and $\vartheta_3=-1.2$. The cubic spline on both parameters leads to non-parallel and non-linear log-quantiles.

3.2. Interpretation of discrete Weibull regression coefficients and output

After a DW regression model has been estimated, the following results can be obtained:

(a) the fitted values for the central trend of the conditional distribution, namely

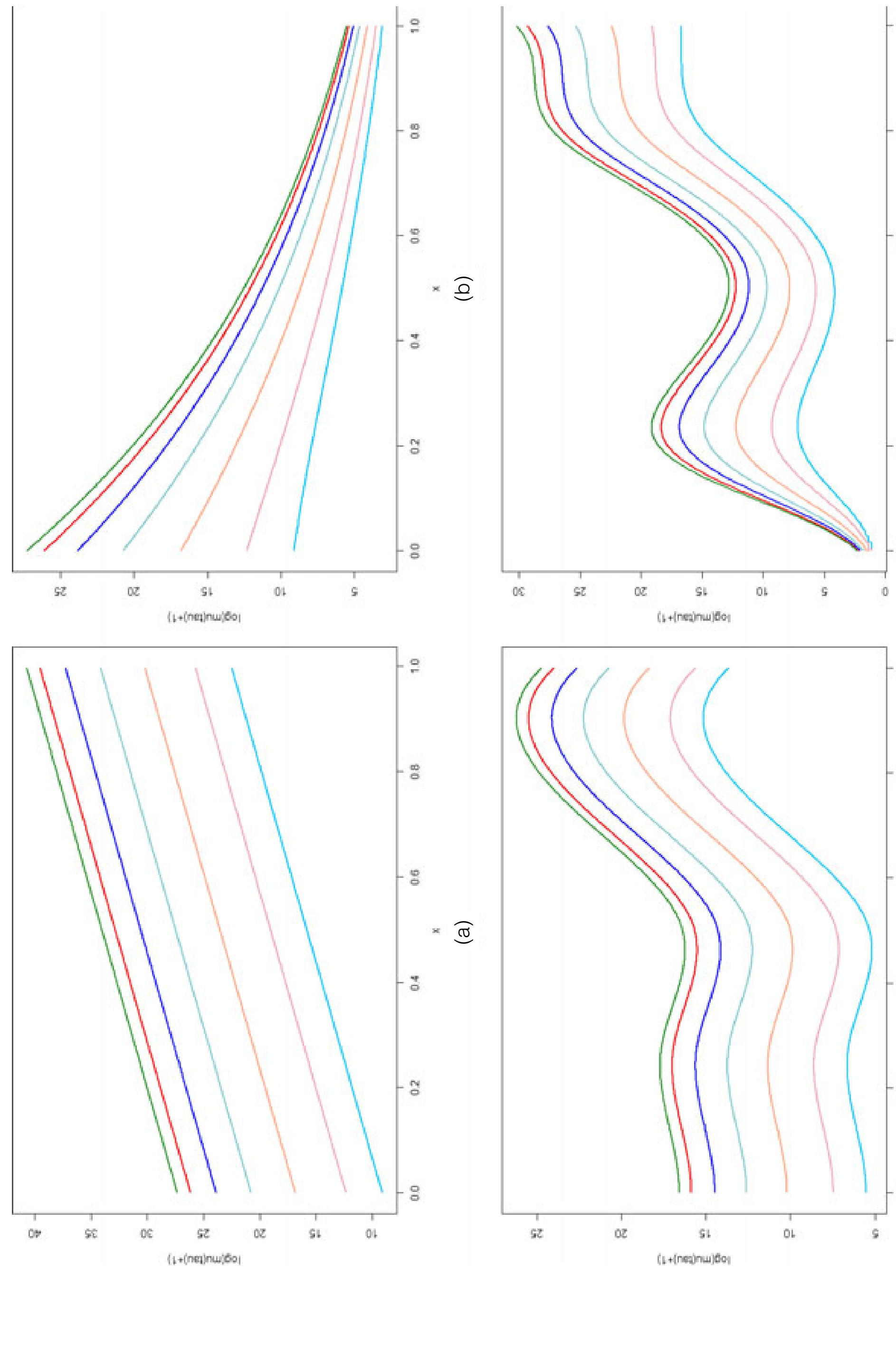


Fig. 2. Conditional quantiles for DW models under (a), (b) linear and (c), (d) non-linear models, and (a), (c) β fixed and (b), (d) not fixed: , $\tau = 0.05$; , $\tau = 0.1$; , $\tau = 0.25$; , $\tau = 0.5$; , $\tau = 0.75$; , $\tau = 0.9$; , $\tau = 0.95$

(i) the mean (equation (2), as mentioned earlier, can be calculated numerically by using the approximated moments of the DW distribution (Barbiero, 2015)) and
(ii) the median (equation (5) can be applied; owing to the skewness, which is common for count data, the median is more appropriate than the mean);

(b) the conditional quantile for any τ by using equation (3), i.e.

$$\mu_{(\tau)}(x) = \lceil \mu_{(*\tau)}(x) \rceil = \left\lceil \left[\frac{\log(1-\tau)}{\log\{q(x)\}} \right]^{1/\beta(x)} - 1 \right\rceil.$$

The analytical expression of the quantiles, combined with the chosen link function, offers a way of interpreting the parameters. Considering a simple regression model on q (case (a) above; Klakattawi *et al.* (2018)), equation (4) at the median leads to

$$\log\{\mu_{(*0.5)}(x) + 1\} = \frac{1}{\beta} \log\{\log(2)\} - \frac{1}{\beta} x' \boldsymbol{\theta}.$$

Thus, the regression parameters $\boldsymbol{\theta}$ can be interpreted in relation to the logarithm of the median, by analogy with Poisson and negative binomial models where the parameters are linked to the logarithm of the mean. In particular, $[\log\{\log(2)\} - \theta_0]/\beta$ is related to the conditional median when all covariates are set to 0, whereas $-\theta_p/\beta$, $p = 1, \ldots, P$, can be related to the change in the median of the response corresponding to a 1-unit change in X_p, keeping all other covariates constant.

For more complex models, partial effects can be computed for each covariate and for each quantile as in Machado and Santos Silva (2005). In particular, let x^0 denote the vector of predictors, where each predictor is set to their sample mean $\bar{x}$ if continuous and to 0 if dummy. Then, the effect for the regressor x_p on the τ-quantile of the response is calculated as the difference

$$\mathrm{PE}(x_p, \tau) = \mu_{(*\tau)}(x_p^1) - \mu_{(*\tau)}(x^0), \tag{7}$$

where x_p^1 is equal to x^0 for all entries with the exception of the pth entry which is increased by 1 unit.

3.3. Model selection and diagnostic checks

Model selection, in terms of polynomial degree and the number of interior knots, is carried out based on known model selection criteria. In this paper we shall use the Akaike information criterion AIC. After fitting a DW regression model, the goodness of fit is checked on the basis of the randomized quantile residuals, as developed by Dunn and Smyth (1996) and advised in the case of non-Gaussian responses (Rigby and Stasinopoulos, 2005). In particular, let

$$\hat{r}_i = \Phi^{-1}(u_i) \qquad i = 1, \ldots, n$$

where $\Phi(\cdot)$ is the standard normal distribution function and u_i is a uniform random variable on the interval

$$\begin{aligned}(a_i, b_i] &= (\lim_{y \uparrow y_i} F(y; \hat{q}_i, \hat{\beta}_i), F(y_i; \hat{q}_i, \hat{\beta}_i)] \\ &\approx [F(y_i - 1; \hat{q}_i, \hat{\beta}_i), F(y_i; \hat{q}_i, \hat{\beta}_i)].\end{aligned}$$

These residuals are expected to follow the standard normal distribution if the model is correct. Hence, the validity of a DW model can be assessed by using goodness-of-fit investigations of the normality of the residuals, such as *Q–Q*-plots and normality tests.

4. Assessing the performance of discrete Weibull regression models

This section performs a comparison of DW regression models with existing parametric approaches and with the jittering approach of Machado and Santos Silva (2005). As the jittering approach fits each conditional quantile separately, we measure the performance of the models for three selected quantiles, namely $\tau = 0.25, 0.5, 0.75$. For each τ and for each model, we evaluate the accuracy in the estimation of the conditional quantile by calculating the root-mean-squared error:

$$\mathrm{RMSE} = \sqrt{\left[\sum_{i=1}^{n} \{\hat{\mu}_{i(\tau)} - \mu_{i(\tau)}\}^2 / n \right]}, \tag{8}$$

where n is the sample size, $\mu_{i(\tau)}$ is the true quantile at observation x_i and $\hat{\mu}_{i(\tau)}$ is the corresponding fitted quantile from the specified model. For DW distributions, this is calculated by using equation (3).

We compare our approach with the following parametric approaches: Poisson, negative binomial, Conway–Maxwell–Poisson, generalized Poisson and Poisson–inverse Gaussian. All distributions, and corresponding generalized additive models, are implemented in the R package `gamlss`, with the exception of the Conway–Maxwell–Poisson model. We therefore opt for an unpenalized basis spline formulation of the generalized additive models and implement this also for the Conway–Maxwell–Poisson model by using the `glm.cmp` function in the R package `COMPoissonReg` (Sellers *et al.*, 2010). Similarly, for the non-parametric jittering approach, we implement a generalized additive model by using the `rq.counts` function in the `Qtools` package (Geraci, 2017).

4.1. Simulation 1: simulating data from our discrete Weibull model

We first simulate data from our proposed model (equation (6)). We consider various sample sizes, $n = 50, 100, 1000$, a covariate $X \sim N(0, 1)$ and a conditional DW distribution where we link the parameters to the covariates by using a generalized additive model under two levels of dispersion. In particular, we generate data by using a cubic B-spline model for $q(x)$ and $\beta(x)$ with the parameters given by

$$\begin{aligned} \log[-\log\{q(x)\}] = & -5 - 5x - 6x^2 - 4x^3 - 8(x - g_1)^3 I(x > g_1) \\ & - 9(x - g_2)^3 I(x > g_2) - 8(x - g_3)^3 I(x > g_3), \end{aligned}$$

(a) (overdispersed)

$$\begin{aligned} \log\{\beta(x)\} = & 0.8 + 0.8x + 0.9x^2 + 0.8x^3 + 1.2(x - g_1)^3 I(x > g_1) \\ & + 1.1(x - g_2)^3 I(x > g_2) + (x - g_3)^3 I(x > g_3), \end{aligned}$$

(b) (underdispersed)

$$\begin{aligned} \log\{\beta(x)\} = & 1.2 + 1.3x + 1.1x^2 + 1.2x^3 + 1.5(x - g_1)^3 I(x > g_1) \\ & + 1.3(x - g_2)^3 I(x > g_2) + (x - g_3)^3 I(x > g_3). \end{aligned}$$

This generally corresponds to the fourth, more complex, case in Fig. 2(d). Setting the values as above leads to overdispersion values between 1.3 and 3.2 and underdispersion values between 0.4 and 0.7.

Fig. 3 reports the errors in equation (8), averaged over 100 iterations and across the three different quantiles ($\tau = 0.25, 0.5, 0.75$). For each model, we consider a B-spline generalized additive

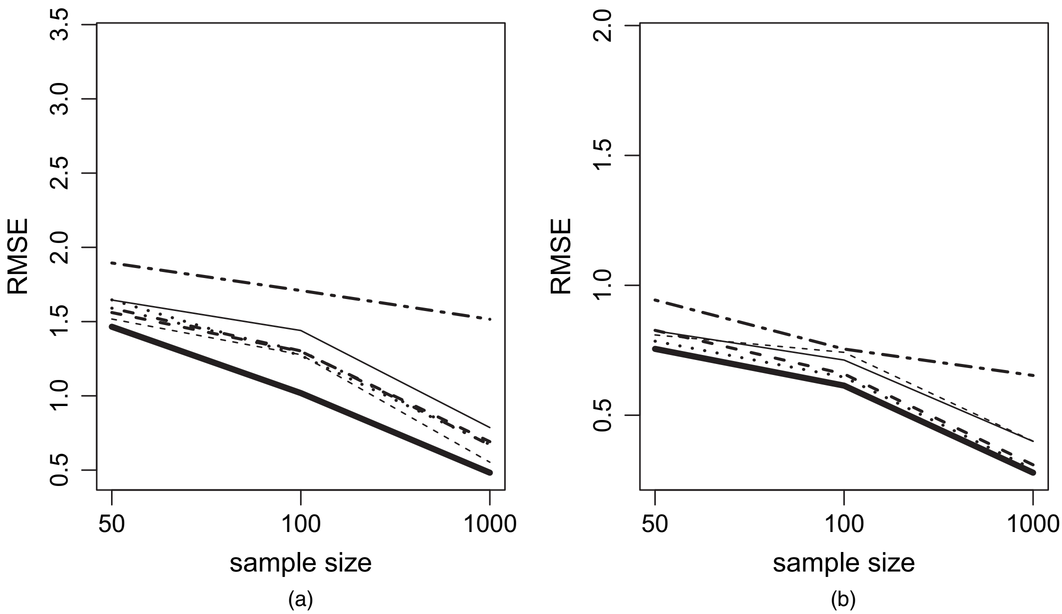


Fig. 3. Comparison of various models in terms of root-mean-squared error (equation (8)) on (a) overdispersed and (b) underdispersed data simulated from a DW model using a cubic *B*-spline on $q(x)$ and $\beta(x)$, with increasing sample sizes ($n=50, 100, 1000$) (the error is averaged over 100 iterations and across three quantiles ($\tau=0.25, 0.5, 0.75$)): ———, DW (*B*-spline); ———, Poisson (*B*-spline); – – –, Conway–Maxwell–Poisson (*B*-spline); - — -, negative binomial (*B*-spline); ------, generalized Poisson (*B*-spline); · · · · · ·, jittering (*B*-spline); - —— -, DW (linear)

Table 1. Simulated data from the negative binomial model of equation (9): comparison of jittering, DW with linear link on q and β, DW with linear link on q and constant β, negative binomial model with linear link on μ and σ†

τ	*Jittering*			$DW\{q(x),\beta\}$			$DW\{q(x),\beta(x)\}$			$NB\{\mu(x),\sigma(x)\}$		
	$n=50$	$n=100$	$n=1000$	$n=50$	$n=100$	$n=1000$	$n=50$	$n=100$	$n=1000$	$n=50$	$n=100$	$n=1000$
0.25	0.607	0.483	0.269	0.782	0.493	0.386	0.568	0.422	0.252	0.526	0.421	0.208
0.50	0.668	0.442	0.262	0.743	0.505	0.305	0.591	0.444	0.250	0.580	0.439	0.219
0.75	0.945	0.603	0.274	0.928	0.618	0.269	0.858	0.576	0.266	0.782	0.543	0.244

†The RMSE is calculated at three different quantiles.

specification with the same complexity as the data-generating model ($D=K=D'=K'=3$) and we place the inner knots at the quartiles of the covariate X. The same is done for the jittering approach. Note, however, that the jittering approach contains many more parameters than the parametric approaches since in this case the generalized additive models are fitted for each quantile. To assess the real need for the generalized additive component, we also include in the simulation a simpler DW regression model with a linear link on q and β ($D=D'=1; K=K'=0$). As expected, the DW *B*-spline model fits better than the other models, whereas the DW linear model cannot compete with the non-linear models. The jittering *B*-spline approach shows good flexibility across the different scenarios, but some instability in parameter estimation was observed at small sample sizes. The ranking of the parametric models is not always consistent. Across various additional experiments and alternative models, we found that the DW model is followed closely by the negative binomial and Poisson–inverse gamma models in the overdispersed case and by the Conway–Maxwell–Poisson model in the underdispersed case. The computational time between the DW and Conway–Maxwell–Poisson models, however, is strikingly different, with the Conway–Maxwell–Poisson model taking approximately 129 s, compared with the DW model which took only 0.3 s, for a single simulation of the results reported here.

4.2. Simulation 2: simulating data from a negative binomial model with tail effects

In a second simulation, we test the performance of our approach in the case of misspecification and tail behaviour. In particular, we simulate data with a negative binomial conditional distribution, with the parameters μ (mean) and σ (dispersion) linked to the covariates by

$$\begin{aligned} \log\{\mu(x)\} &= 0.6 + 0.4x_1, \\ \log\{\sigma(x)\} &= -5 + 2x_2. \end{aligned} \tag{9}$$

We simulate tail behaviour by letting the dispersion parameter depend on a regressor that does not affect the mean.

Table 1 reports the square root of the error in equation (8), averaged over 100 iterations, with X_1 and X_2 drawn from an $N(0,1)$ distribution and for three different sample sizes ($n = 50, 100, 1000$). The complexity of the models is selected to match that of the generating model (equation (9)), i.e. a linear link on both parameters. For the DW model, we also considered a simpler model where β is held constant. It is interesting that the more complex DW model is performing similarly to the well-specified negative binomial model, whereas keeping β constant in the DW model leads to poorer results. Indeed, looking at the parameter estimates of the more complex DW model, X_1 is found significant in predicting q whereas X_2 is significant in the regression on β, suggesting that the regression on both parameters helps to identify more complex dependences such as tail behaviour.

5. Modelling the relationship between family background and planned fertility

We use the latest data from the national survey of demographic dynamics in Mexico (Instituto Nacional de Estadístíca y Geografía, 2014) to study the effect of education and family background on fertility plans. In particular, we take as dependent variable the number of planned children declared by young Mexican women at the survey interview. As in Miranda (2008), we consider women between 15 and 17 years old who at the time of the interview were living with at least one biological parent and had neither started independent economic life nor entered motherhood. This selection avoids any confusion between current and planned fertility and ensures that all individuals are broadly at the same point of their life cycle. Some covariates are selected to control for education and family background: whether the teenager can speak an indigenous or native language, whether primary, secondary and higher education attainments were completed, a wealth index (low, medium low, medium high and high), the location of the parental household (rural, urban and suburban) and a set of variables describing the socio-economic characteristics of the head of the family, among which the age and gender and the same educational attainment covariates were considered for the teenagers. Finally, the total number of people living with the teenager (family size) and a series of dummies indicating the state of residence (32 in total) are also used as explanatory variables. This gives a total of 53 explanatory variables in this study and a sample size of 5906 women. Fig. 4 shows the distribution of the

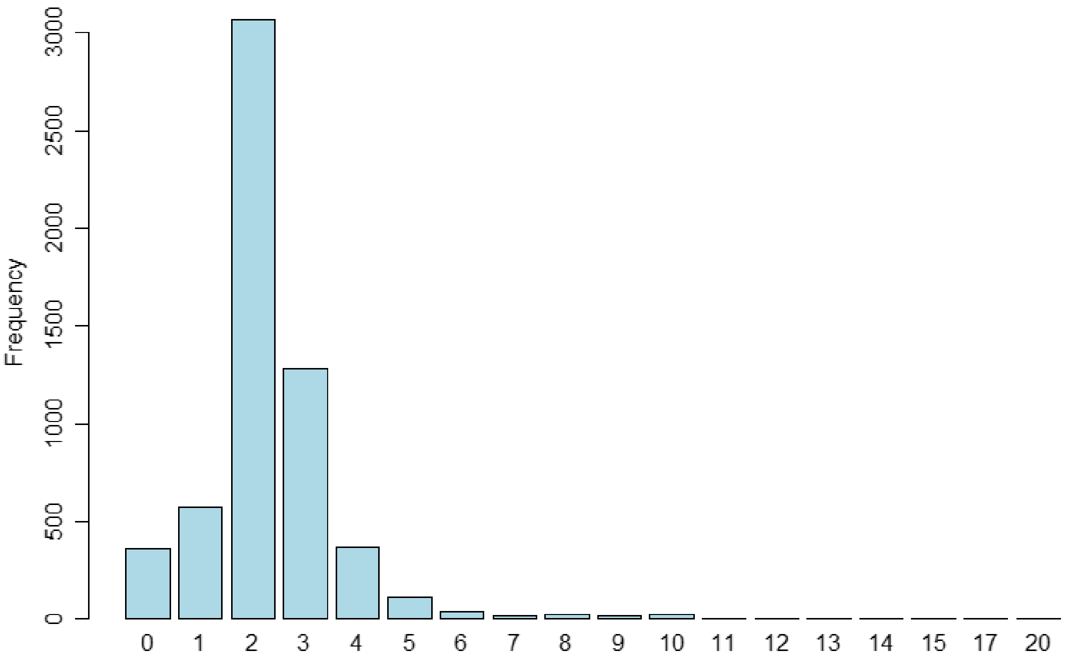


Fig. 4. Distribution of the response variable: ideal number of children for 15–17-years-old Mexican women

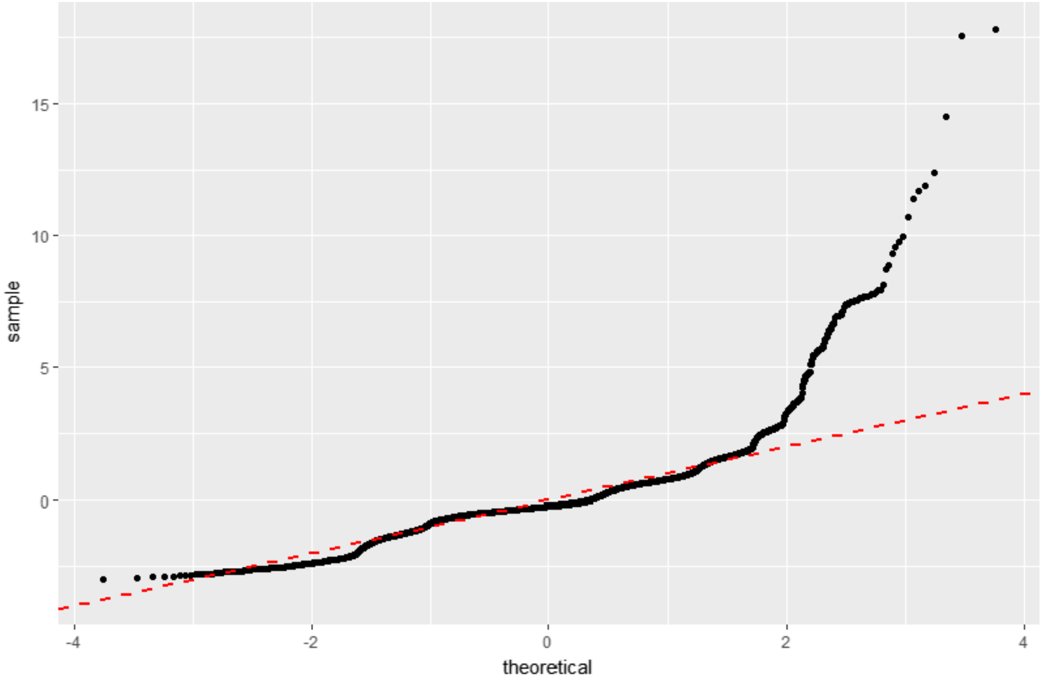
Table 2. The Poisson, DW and Conway–Maxwell–Poisson generalized additive models selected on the ideal fertility data with a *B*-spline (with degree *D* and number of internal knots *k*) for two of the variables†

Model	*AIC*	*Log-likelihood*	*Number of parameters*	*Age (family head)*		*Family size*		*Time (s)*
				D	*k*	*D*	*k*	
Poission	19779.59	−9838.796	51	1	1	1	0	0.14
DW	19351.46	−9561.731	114	2	3	2	2	9.67
Conway–Maxwell–Poisson	19175.24	−9475.621	112	1	1	3	3	7339.46

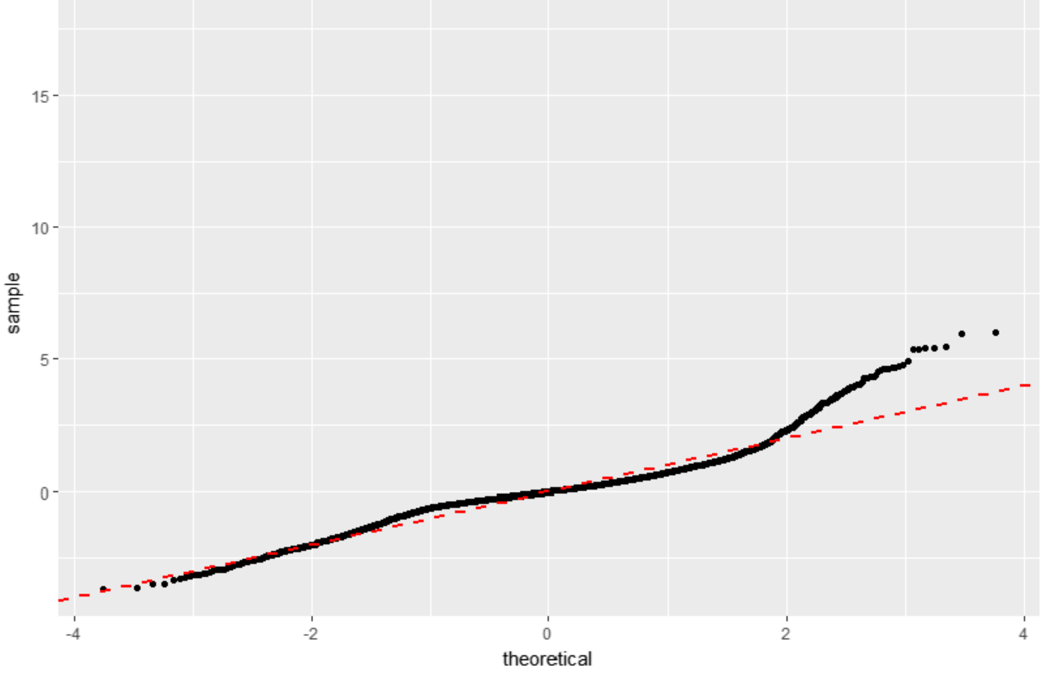
†The last column reports the system time.

response variable. Without taking into consideration the effect of the covariates, the distribution of ideal fertility shows underdispersion relative to a Poisson distribution with a dispersion index of 0.86, higher than that of the 1997 data (0.55), possibly because a small number of outliers (eight women) declared more than 12 planned children (the maximum in Miranda (2008)).

We fit Poisson, Conway–Maxwell–Poisson and DW generalized additive models to these data (the generalized Poisson model had problems of convergence with this data set). As penalized regression splines are not implemented for Conway–Maxwell–Poisson models, unpenalized *B*-splines were used for all models, with knots placed at the quantiles of the covariates. Given the computational complexity of the Conway–Maxwell–Poisson model but in the interest of a fair comparison, we select the best model with the following strategy: we fix a linear link on all parameters (one for the Poisson and two for the DW and Conway–Maxwell–Poisson models);



(a)



(b)

Fig. 5. *Q–Q*-plot of the randomized quantile residuals of the (a) Conway–Maxwell–Poisson and (b) DW generalized additive models fitted to the ideal fertility data

Table 3. Partial effects on ideal fertility for the jittering (top) and the DW (bottom) generalized additive models for the regressors that are found significant at the 5% level by the DW model†

Regressor	*Results for the following values of τ:*					
	0.25	*0.5*	*0.75*	*0.9*	*0.95*	*0.99*
Jittering						
family size	−0.01	*−0.038*	*−0.054*	−0.069	−0.076	*−0.362*
HFage	0.037	*0.035*	*0.041*	0.12	0.168	0.317
cprimary	*6.971*	*2.929*	2.216	*1.221*	*1.624*	5.082
isecondary	*7.233*	*2.729*	1.942	*0.89*	*1.447*	5.443
csecondary	*7.336*	*2.82*	2.047	0.706	1.837	3.016
osecondary	*7.335*	*2.811*	2.019	*0.985*	*1.481*	5.549
wealth_mlow	*−0.077*	*−0.111*	*−0.167*	*−0.166*	0.109	0.564
wealth_mhigh	*−0.074*	*−0.128*	*−0.183*	−0.096	0.245	*2.477*
wealth_high	−0.087	−0.07	−0.063	0.052	0.979	*3.852*
HFmale	0.048	0.011	0.053	0.039	−0.044	−1.062
HFcprimary	0.111	0.056	0.061	−0.117	−0.218	−1.664
HFisecondary	0.101	0.029	0.021	−0.185	−0.338	−1.151
HFcsecondary	−0.031	0.09	0.178	−0.039	−0.324	−2.209
HFosecondary	0.062	0.022	0.085	0.06	0.163	0.031
DW						
family size	−0.003	−0.042	−0.096	−0.155	−0.194	−0.273
HFage	0.028	0.045	0.066	0.087	0.1	0.126
cprimary	2.386	2.625	2.741	2.756	2.731	2.627
isecondary	2.429	2.582	2.598	2.517	2.434	2.220
csecondary	2.814	2.79	2.607	2.337	2.139	1.710
osecondary	2.517	2.66	2.663	2.568	2.475	2.243
wealth_mlow	−0.265	−0.179	−0.005	0.223	0.387	0.746
wealth_mhigh	−0.405	−0.215	0.172	0.684	1.059	1.888
wealth_high	−0.344	−0.094	0.377	0.984	1.423	2.387
HFmale	0.095	0.033	−0.064	−0.176	−0.252	−0.409
HFcprimary	0.313	0.134	−0.129	−0.422	−0.618	−1.017
HFisecondary	0.221	0.083	−0.124	−0.357	−0.513	−0.833
HFcsecondary	0.266	0.063	−0.227	−0.546	−0.757	−1.184
HFosecondary	0.133	0.09	0.014	−0.077	−0.141	−0.274

†For jittering, significance at 5% for each quantile is highlighted in italics.

then we consider the inclusion of possible non-linear terms for the two non-categorical variables (age of the head of the family and family size) in a forward stepwise manner. For this, we select the same level of complexity for each parameter and we search for all model combinations up to a maximum of degree 3 of the polynomial and three internal knots (a cubic spline). The best model at each step is selected on the basis of AIC. Table 2 shows that the model selected resulted in a quadratic spline for the two continuous variables in this study. Fig. 5 shows the fitting of the top two models in terms of randomized quantile residuals. Overall, the DW and Conway–Maxwell–Poisson models appear to provide a similar fit to the data, with a slightly lower AIC for the Conway–Maxwell–Poisson model but a worse fit of the Gaussian distribution to the residuals. The computational time, here reported only for the best model in the search, shows a striking difference between the DW and Conway–Maxwell–Poisson models, which limits further comparisons and more extensive searches. We have further compared the selected DW model with a penalized spline model, using the same degree and number of (equally spaced) knots as

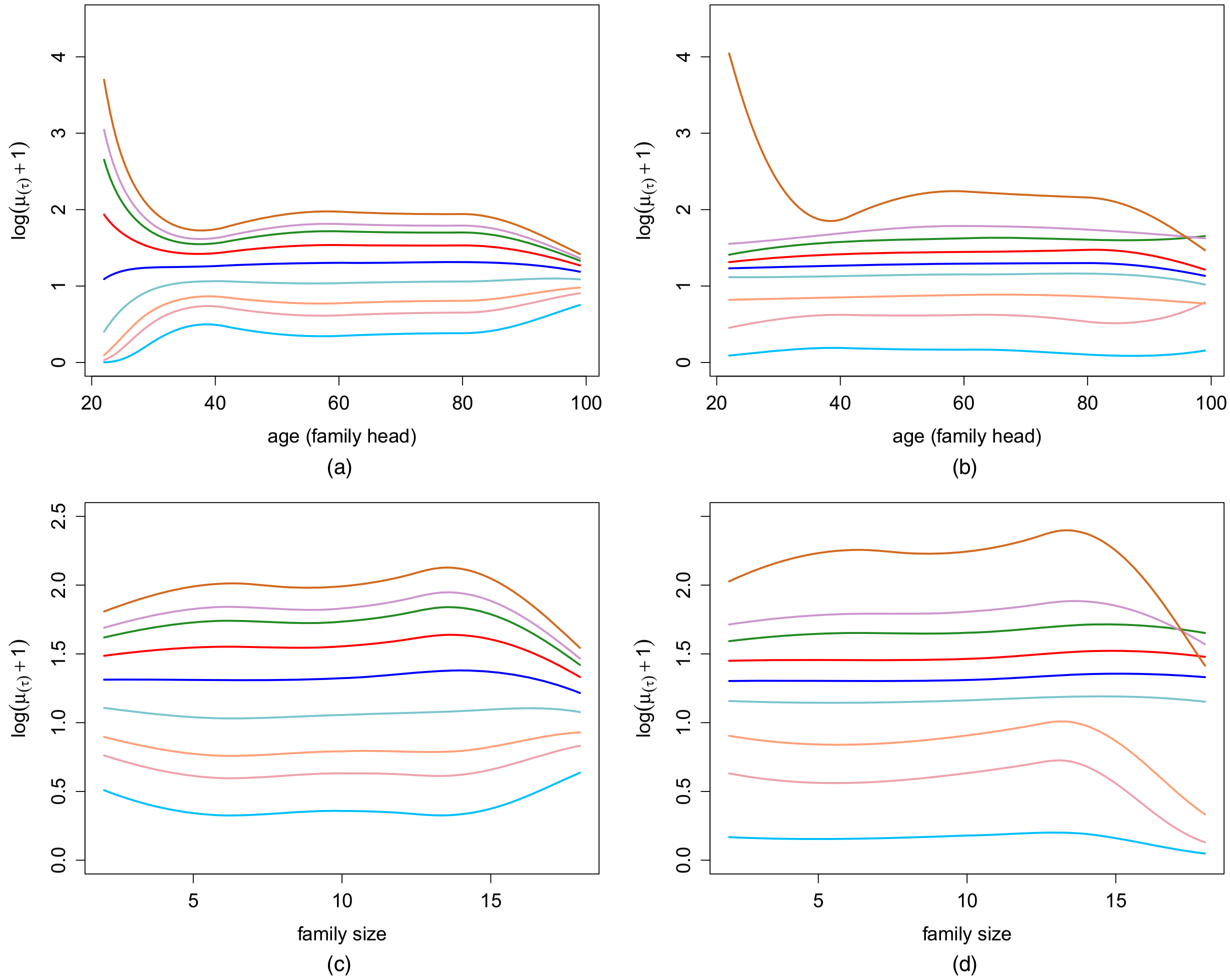


Fig. 6. Conditional τ-quantiles of ideal fertility for the two variables (a), (b) age of the family head and (c), (d) family size, keeping all the other covariates fixed to their mean, for (a), (c) the DW generalized additive model and (b), (d) the jittering approach: ——, $\tau = 0.01$; ——, $\tau = 0.05$; ——, $\tau = 0.1$; ——, $\tau = 0.25$; ——, $\tau = 0.5$; ——, $\tau = 0.75$; ——, $\tau = 0.9$; ——, $\tau = 0.95$; ——, $\tau = 0.99$

the selected model and imposing a penalty on the third-order differences of the parameters (AIC $= 19374.33$), and with a smoothing cubic spline (AIC $= 19373.38$). Overall, no improvement was observed over the unpenalized B-spline model of Table 2, but all models were superior to a simple DW model with a linear link on q and β (AIC $= 19410.49$). Finally, using this simple linear model as the baseline model, the more complex B-spline models did not show signs of variance inflation, with generalized variance inflation factors close to 1 in all cases, e.g. the Cox and Snell and the Cragg and Uhler generalized R^2-factors are 1.0148 and 1.0155 respectively for the unpenalized B-spline model (Nagelkerke, 1991).

To summarize the findings from our fitted model and in an attempt to measure how well a flexible parametric approach can approximate the conditional distribution of ideal fertility given the explanatory variables, we compare the partial effects that were obtained from our model with those of a more flexible non-parametric jittering approach where a quadratic spline (with the same number of internal knots as in the DW model) is fitted to each conditional quantile, thus resulting in a larger number of parameters. The effects are calculated as in equation (7) and are reported in Table 3 only for the variables that are found significant at the 5% level by the DW model. Table 3 shows similar levels of effects between the two approaches, in terms of both sign and intensity. In addition, both approaches can capture tail effects in the distribution, with several variables exhibiting sign reversals of the effects. The conclusions are similar to those obtained by Miranda (2008) in their earlier study, with variables related to education and family background being highly significant. In our analysis, and also using the Conway–Maxwell–Poisson model, the education indicators of the head of the family appear to be highly significant whereas these are not picked up as significant by the jittering approach although the marginal effects are close. This may be down to a higher instability in the estimation of the standard errors for jittering, possibly due to the larger number of parameters and also to the uniform random sampling underlying the method (100 samples are used for the results that are presented in this paper). A further example of this is with the variable family size, which is not found significant for the Conway–Maxwell–Poisson model (p-values above 0.1) but is found highly significant for the DW model and only for some of the quantiles for the jittering approach (e.g. 0.99 but not 0.9 and 0.95). Although detecting τ-dependent significant variables can be seen as an advantage of the jittering approach over the parametric approaches, this can be seen as a limitation when the conclusions are overly sensitive to the specific quantile selected, as observed in this instance. Fig. 6 shows further how this discontinuity could be the result of crossing of quantiles, which is a further drawback of fitting models individually for each quantile. Although the general trends are similar between DW and jittering, the jittering approach produces crossing of quantiles at the extreme of the distribution, where there is usually a small sample size.

6. Conclusions

Motivated by an investigation about the dependence of planned fertility on education and family background, we develop a novel regression model for count data based on the DW distribution, which has had limited use to date. We show how a regression model based on this distribution can provide a simple and unified framework to capture different levels of dispersion in the data, namely underdispersion and overdispersion. Given the expected complex dependences in the planned fertility study, we develop a generalized additive model to link both parameters of the distribution to the explanatory variables.

Through a simulation study and the real data application we show some important features of the models proposed, which could potentially lead to their wide applicability to the modelling of count data. Firstly, the conditional quantiles have a simple analytical formula, which makes the calculation of partial effects straightforward as well as the interpretation of the regression coefficients. Secondly, the likelihood is the same as that of a continuous Weibull distribution with interval-censored data, so efficient implementations are already available in the R package `gamlss`, for a range of models, including mixed and mixture models, and inferential procedures, including penalized likelihood approaches. Thirdly, the distribution can capture both cases of overdispersion and underdispersion, similarly to the Conway–Maxwell–Poisson distribution but at a fraction of the time. Fourthly, a generalized additive DW model can compete against the more flexible quantile regression approaches, without the need of individual fitting at each quantile and without the inherent issue of conditional quantiles crossing.

The analysis on the planned fertility study showed overall a good fit of the model proposed compared with the existing parametric methods and similar results to those obtained by the non-parametric jittering approach. Jittering, however, suffered from some instability in the estimation of the parameters at the extreme quantiles and crossing of the conditionals' quantile. A goodness-of-fit check of the randomized quantile residuals from the DW model selected showed some departure from normality at the tails of the distribution. Future work will extend the approach proposed in this paper to more flexible DW distributions, such as the three-parameter Burr type XII distribution (Para and Jan, 2016) which was found to provide a good fit to the ideal fertility response variable (AIC(Burr) = 18400.36 *versus* AIC(DW) = 20047.40) but which has not been implemented in a regression context.

Acknowledgements

We thank Professor Joao Santos Silva for useful discussions and insights on this paper. This work has been supported in part by a National Institute for Health Research method grant (NIHR-RMOFS-2013-0309).

References

Barbiero, A. (2015) DiscreteWeibull: discrete Weibull distributions (type 1 and 3). *R Package Version 1.0.1*. Università degli Studi di Milano, Milan. (Available from `http://CRAN.R-project.org/package=DiscreteWeibull`.)

Cameron, A. C. and Trivedi, P. K. (2013) *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Chakraborty, S. (2015) Generating discrete analogues of continuous probability distributions—a survey of methods and constructions. *J. Statist. Distribns Appl.*, **2**, 1–30.

Chanialidis, C., Evers, L., Neocleous, T. and Nobile, A. (2018) Efficient Bayesian inference for COM-Poisson regression models. *Statist. Comput.*, **28**, 595–608.

Consul, P. and Famoye, F. (1992) Generalized Poisson regression model. *Communs Statist. Theory Meth.*, **21**, 89–109.

De Boor, C. (1972) On calculating with B-splines. *J. Approximn Theory*, **6**, 50–62.

Dierckx, P. (1995) *Curve and Surface Fitting with Splines*. Oxford: Oxford University Press.

Dunn, P. K. and Smyth, G. K. (1996) Randomized quantile residuals. *J. Computnl Graph. Statist.*, **5**, 236–244.

Eilers, P. H. C. and Marx, B. D. (1996) Flexible smoothing with *B*-splines and penalties. *Statist. Sci.*, **11**, 89–121.

Englehardt, J. D., Ashbolt, N. J., Loewenstine, C., Gadzinski, E. R. and Ayenu-Prah, A. Y. (2012) Methods for assessing long-term mean pathogen count in drinking water and risk management implications. *J. Wat. Hlth*, **10**, 197–208.

Englehardt, J. D. and Li, R. (2011) The discrete Weibull distribution: an alternative for correlated counts with confirmation for microbial counts in water. *Risk Anal.*, **31**, 370–381.

Geraci, M. (2017) Qtools: utilities for quantiles. *R Package Version 1.2*. University of South Carolina, Columbia. (Available from `https://cran.r-project.org/package=Qtools`.)

Haselimashhadi, H., Vinciotti, V. and Yu, K. (2018) A novel Bayesian regression model for counts with an application to health data. *J. Appl. Statist.*, **45**, 1085–1105.

Hilbe, J. M. (2014) *Modeling Count Data*. Cambridge: Cambridge University Press.

Horowitz, J. (1992) A smooth binary score estimator for the binary response model. *Econometrica*, **60**, 505–531.
Instituto Nacional de Estadística y Geografía (2014) National survey of demographic dynamics, 2014. Instituto Nacional de Estadística y Geografía, Mexico City.
Kadane, J. B., Krishnan, R. and Shmueli, G. (2006) A data disclosure policy for count data based on the COM-Poisson distribution. *Mangmnt Sci.*, **52**, 1610–1617.
Khan, M. A., Khalique, A. and Abouammoh, A. (1989) On estimating parameters in a discrete Weibull distribution. *IEEE Trans. Reliab.*, **38**, 348–350.
Klakattawi, H., Vinciotti, V. and Yu, K. (2018) A simple and adaptive dispersion regression model for count data. *Entropy*, **20**, 142.
Knodel, J. and Prachuabmoh, V. (1973) Desired family size in Thailand: are the responses meaningful? *Demography*, **10,** 619–637.
Kulasekera, K. (1994) Approximate MLEs of the parameters of a discrete Weibull distribution with type I censored data. *Micrelectron. Reliab.*, **34**, 1185–1188.
Lee, M. (1992) Median regression for ordered discrete response. *J. Econmetr.*, **51**, 59–77.
Lee, D. and Neocleous, T. (2010) Bayesian quantile regression for count data with application to environmental epidemiology. *Appl. Statist.*, **59**, 905–920.
Machado, J. and Santos Silva, M. (2005) Quantiles for counts. *J. Am. Statist. Ass.*, **100**, 1226–1237.
Manski, C. (1985) Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *J. Econmetr.*, **3**, 205–228.
Miranda, A. (2008) Planned fertility and family background: a quantile regression for counts analysis. *J. Popln Econ.*, **21**, 67–81.
Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.
Nakagawa, T. and Osaki, S. (1975) The discrete Weibull distribution. *IEEE Trans. Reliab.*, **24**, 300–301.
Nelder, J. A. and Wedderburn, R. W. (1972) Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.
Noufaily, A. and Jones, M. (2013) Parametric quantile regression based on the generalized gamma distribution. *Appl. Statist*., **62**, 723–740.
Para, B. and Jan, T. (2016) On discrete three parameter Burr Type XII and discrete Lomax distributions and their applications to model count data from medical science. *Biometr. Biostatist. Int. J.*, **4**, article 00092.
Pritchett, L. H. (1994) Desired fertility and the impact of population policies. *Popln Develpmnt Rev.*, **20**, 1–55.
Rigby, R. A. and Stasinopoulos, D. M. (2005) Generalized additive models for location, scale and shape (with discussion). *Appl. Statist.*, **54**, 507–554.
Sáez-Castillo, A. and Conde-Sánchez, A. (2013) A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computnl Statist. Data Anal.*, **61**, 148–157.
Sellers, K., Lotze, T. and Raim, A. (2010) COMPoissonReg: COM-Poisson and zero-inflated COM-Poisson regression. *R Package Version 0.4.1.* Georgetown University, Washington DC. (Available from `https://CRAN.R-project.org/package=COMPoissonReg`.)
Sellers, K. F. and Shmueli, G. (2010) A flexible regression model for count data. *Ann. Appl. Statist.*, **4**, 943–961.
Smith, D. and Faddy, M. (2016) Mean and variance modeling of under- and overdispersed count data. *J. Statist. Softwr.*, **69**, 1–23.
Stasinopoulos, D. M. and Rigby, R. A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *J. Statist. Softwr.*, **23**, 1–46.
Stasinopoulos, D. M., Rigby, R. A., De Bastiani, F. and Voudouris, V. (2017) *Flexible Regression and Smoothing: Using GAMLSS in R*. Boca Raton: Chapman and Hall–CRC.
Willmot, G. E. (1987) The Poisson-inverse Gaussian distribution as an alternative to the negative Binomial. *Scand. Act. J.*, 113–127.